QUANTITATIVE ANALYSIS OF BIOLOGICAL DECISION SWITCHES

A Dissertation
Presented to
The Academic Faculty

by

In-Ho Joh

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology
May 2011

QUANTITATIVE ANALYSIS OF BIOLOGICAL DECISION SWITCHES

Approved by:

Professor Joshua S. Weitz, Advisor
School of Biology
Georgia Institute of Technology

Professor Eberhard O. Voit
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Professor Harold Kim
School of Physics
Georgia Institute of Technology

Professor Roger Wartell
School of Biology
Georgia Institute of Technology

Professor Kurt Wiesenfeld
School of Physics
Georgia Institute of Technology

Date Approved: Mar 28 2011

To Jiyoung and Lydia

ACKNOWLEDGEMENTS

My study would not be successful without many people who supported me. I would like to thank Professor Joshua Weitz for his invaluable advices throughout my PhD studies. His guidance was a critical element in making my studies successful. I am also grateful to my thesis committee members, Professor Kurt Wiesenfeld, Harold Kim, Eberhard Voit and Roger Wartell, for insightful comments.

All former and current members of Weitz lab were integral to my PhD studies. Especially, Dr. Yuriy Mileyko, Hao Wang and Taejun Lee have offered me invaluable assistance. Special thanks to Gabriel Mitchell with whom I had numerous thought provoking discussions. Also, Dr. Russell Monds, Anuradha Villapakam and Selwyn Quan at Stanford University guided me into the world of experimental molecular biology. I also thank Dr. Lanying Zeng at University of Illinois at Urbana-Champaign and Professor Ido Golding at Baylor College of Medicine for sharing the results from single cell experiments with me.

I deeply appreciate all the support from my family and friends. Foremost, my deepest gratitude goes to my wife, Jiyoung, and daughter, Lydia, for their endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Biological organisms from multicellular eukaryotes to bacteria and even viruses often exhibit multiple alternative phenotypes or behaviors to adapt to various environmental stimuli. Therefore, an organism chooses one phenotype among multiple alternatives, and we call this process a decision switch. Here we focus on decision switches at the cellular level, and investigate decision switches mediated by gene regulation. Even if environmental stimuli provide inputs for gene regulation, it is intrinsically stochastic, and an isogenic population may exhibit different phenotypes even under almost identical conditions. We focus on two particular aspects of biological decision switches at the cellular level: (1) how changes in copy numbers of genetic components affect gene expression and facilitate multiple alternative phenotypes and (2) how temporal dynamics of gene regulation mediate alternative cell fates and lead to future cell fates.

To study the effect of copy number of genetic components on gene expression, we present a quantitative model of the phage λ decision switch based on models of gene regulatory dynamics. By treating copy numbers of viral genes as variables, we show that the decision between lysis and lysogeny can sensitively depend on the number of coinfecting phages. Therefore, our results suggest that even viruses can adapt their behaviors collectively by sharing their gene products within a host cell. We also show that strong nonlinearity within the gene regulatory networks can lead to bistability, hence facilitate alternative phenotypes. Then we also model other commonly observed genetic circuits while systematically varying copy numbers of genetic circuits. Analyses of these genetic circuits demonstrate that the gene copy

number is an omnipresent parameter that can facilitate alternative phenotypes, and small changes in copy number may lead to drastic changes in gene expression. Further, changes in the copy number of a three-gene motif with successive inhibition can switch between oscillatory and stationary dynamics. In all cases, the qualitative change in gene expression is due to the nonlinear nature of transcriptional feedback.

To investigate how alternative cell fates are mediated by temporal dynamics of gene regulation, we revisit the lysis-lysogeny decision switch of bacteriophages in which cell fates are determined by temporal dynamics of gene regulation unlike the earlier model. We find that increasing the number of coinfecting phages increases the chance of quiescent viral growth, in agreement with prior experimental studies. Predicted heterogeneity in cell fates is shown to agree with experimental data when including a previously unidentified gene dosage compensation mechanism, which represents an alternative hypothesis to explain how multiple phages interact in influencing cell fate. Next we study how cell fates can be predicted given temporal dynamics of gene expression. To do so, we present a quantitative measure of cell fate predictability. Our analyses of simple model systems suggest temporal dynamics of gene expression can be highly correlated with eventual cell fates. Thus, our study quantifies when and how the current state of a cell may serve as an indicator of its future.

CHAPTER I

INTRODUCTION

Organisms adapt to various environmental conditions in order to survive by switching phenotypes or behaviors. Often there are many possible phenotypes, and organisms choose one particular phenotype among many alternatives. This is not only the case for multicellular organisms, but also for bacteria and even viruses. Here we focus on such phenotypic changes at the cellular level, and we refer to them as decision switches. Decision switches can be driven deterministically or stochastically [102] because phenotypic changes are mediated by gene expression within cells. Environmental conditions drive gene expression [194], but biochemical reactions in gene regulatory networks are intrinsically stochastic [53, 183]. Therefore, genetically identical cells often exhibit diverse phenotypes even under almost identical environmental conditions. For example, phage λ, a virus which infect *E. coli* strains, is one of the simplest organisms that exhibit alternative developmental pathway [146]. Upon infecting a host, phages may enter the lytic pathway where phages lyse the host to burst out daughter phages. Or the phages enter the lysogenic pathway where they integrate viral genomes into the host chromosome. Earlier experiments showed that the fraction of lysogeny depends on the number of coinfecting phages, and multiple coinfections predominantly lead to lysogeny [101, 97]. How the bias between lysis and lysogeny is facilitated is largely an open question. Lysis yields fast growth of phages, but lysogeny might be advantageous when the host density becomes low [173].

There is huge interest in understanding cellular decision switches. Estimating risk factors of diseases is an important step towards prevention and treatment especially for cancers and cardiovascular diseases [32, 197, 152]. In development, understanding

how stem cells diverge into multiple lineage is a fundamental question [167, 91], and tight regulation of both gene expression and timing is critical. Also, different phenotypes lead to different fitnesses in various environmental conditions and may directly affect population dynamics [60, 122]. From the evolutionary perspective, multiple alternative phenotypes do not always provide obvious benefit to individuals [173, 12], but might be beneficial as a bet hedging strategy for changing environmental conditions. Here we focus on decision switches within prokaryotes mediated by gene regulatory networks (GRNs) and approach them with two specific emphases: how changes in copy numbers of genes and promoters alter gene expression and how alternative cell fates are mediated by temporal dynamics of gene regulation.

1.1 Copy number variation and gene expression

Copy number variation is a phenomenon found in almost all organisms from higher and unicellular eukaryotes [142, 84, 92] to bacteria [6]. There are many examples which show the resulting gene expression is tightly controlled even with variation in copy number [55, 78]. However, the production rate of a protein is expected to be linearly proportional to the copy number without any feedback mechanisms. Here we systematically change copy numbers of genes and promoters within models of GRNs, and investigate how copy numbers shape gene expression pattern. In doing so, we report that nonlinearity due to transcriptional feedback is essential for copy-number-sensitive gene expression, and may lead to bistability which facilitates alternative phenotypes.

In Chapter 2 [193], we present a quantitative model of how bacterial viruses can make collective decisions concerning the fate of infected cells. For many bacterial viruses, the choice of whether to kill host cells or enter a latent state depends on the multiplicity of coinfection. We base our theory on mechanistic models of gene regulatory dynamics, and unlike most previous works treat the copy number of viral

genes as a variable. When viral regulation of cell fate includes nonlinear feedback loops, very small changes in transcriptional rates can lead to dramatic changes in steady-state gene expression. Hence, we show that decisions can be reached, e.g., lysis or lysogeny, depending on the cellular multiplicity of infection within a broad class of gene regulatory models of viral decision-making. Comparisons of a parameterized version of the model with molecular studies of the decision structure in the temperate bacteriophage λ are consistent with our conclusions. Because the model is generally applicable to other viral systems, it suggests that bacterial viruses can respond adaptively to changes in population dynamics, and the features of collective decision-making in viruses are evolvable life history traits.

In Chapter 3 [124], we extend our framework into other commonly observed genetic circuits which are known as network motifs and demonstrate that the copy number is an omnipresent parameter that can dramatically modify dynamical functions of network motifs. We consider positive feedback, bistable feedback and toggle switch motifs, and show that variation in gene copy number, on the order of a single or few copies, can lead to multiple orders of magnitude change in gene expression. Further, small changes in gene copy number for a three-gene motif with successive inhibition, the repressilator, can lead to a qualitative switch in system behavior among oscillatory and equilibrium dynamics. In all cases, the qualitative change in expression is due to the nonlinear nature of transcriptional feedback in which duplicated motifs interact via common pools of transcription factors. We discuss the relevance of our findings to ongoing efforts to link copy number variation with cell fate determination by viruses, dynamics of synthetic gene circuits, and constraints on evolutionary adaptation.

1.2 Cell fates mediated by temporal dynamics of gene regulation

Often cell fate determination is described as the result of gene regulatory dynamics at equilibrium, but there are many examples showing that changing cell fates occur

within a short time scale. This suggests that decision switches need not be governed by the dynamics at equilibrium, but mediated by temporal dynamics [69]. Temporal dynamics are highly susceptible to noise [53, 177] which makes prediction of future cell fates difficult. Recent development of nondestructive biomarkers enabled observing the change in gene expression almost in real time [190, 147], and temporal dynamics may carry information on cell fates that is unavailable from stationary states [28]. Here we demonstrate that the decision of lysis and lysogeny in phage-λ-infected bacteria can be determined by temporal dynamics and present a quantitative measure for cell fate predictability.

In Chapter 4 [87], we present a quantitative model of alternative cell fates mediated by transient gene regulatory dynamics. Cell fate determination is usually described as the result of the stochastic dynamics of GRNs reaching one of multiple steady-states each of which corresponds to a specific fate. However, the fate of a cell is determined in finite time suggesting the importance of transient dynamics in cellular decision making. Here we consider cellular decision making as resulting from first passage processes of regulatory proteins and examine the effect of transient dynamics within the initial lysis-lysogeny switch of phage λ. Importantly, the fate of an infected cell depends, in part, on the number of coinfecting phages. Using a quantitative model of the phage λ GRN, we find that changes in the likelihood of lysis and lysogeny can be driven by changes in phage coinfection number regardless of whether or not there exists steady-state bistability within the GRN. Furthermore, two GRNs which yield qualitatively distinct steady state behaviors as functions of phage coinfection number can show similar transient responses, sufficient for alternative cell fate determination. We compare our model results to a recent experimental study of cell fate determination in single cell assay of multiply infected bacteria. Whereas the experimental study proposed a quasi-independent hypothesis for cell fate determination consistent with an observed data collapse, we demonstrate that observed cell fate results are

compatible with an alternative form of data collapse consistent with a partial gene dosage compensation mechanism. We show that including partial gene dosage compensation at the mRNA level in our stochastic model of fate determination leads to the same data collapse observed in the single cell study.

In Chapter 5 [86], we propose a quantitative framework for cell fate prediction based on analysis of temporal dynamics of gene regulation. It is well known that cells undergo changes in gene expression in response to environmental stimuli that result in a marked phenotypic change. However, predicting phenotypic changes of cells remains challenging due to noise associated with the dynamics and difficulty in measuring internal cell fates. Prior works suggest that early changes in gene expression may prime the cell for a specific phenotypic change. Hence, detection of correlations between gene expression dynamics and cell fate outcomes can confer dramatic improvements in cell fate prediction. To illustrate how temporal dynamics facilitate cell fate prediction, we develop a statistical approach to quantify the change in predictability over time. We apply it to two model systems: a random walk model and a positive feedback loop. Our analyses of these systems suggest an increase in predictability over time but only when fluctuations in gene expression do not override the intrinsic bias. We apply our statistical approach to experimental data and model results of bacteria infected by phages. We observe an increase in predictability over time for both the experimental data and simulations from a mechanistic model. We also observe decreases with time in the number of cells that have not yet reached a new cell fate. This suggests a tradeoff between gain of predictability and and its scope. Optimization of criteria for predicting cell fate may facilitate improvements of experimental strategies when there are benefits and costs associated with measuring gene expression.

CHAPTER II

COLLECTIVE DECISION MAKING IN BACTERIAL VIRUSES[1]

2.1 Introduction

Bacterial viruses ("bacteriophages" or "phages") can be classified based on their life histories into two broad categories: virulent and temperate [48, 26]. Virulent phages possess two life history stages, an extracellular stage, in which a metabolically inactive virion passively diffuses in the environment, and an intracellular stage, in which the viral genome redirects transcription and translation leading to virion production and cell lysis. In contrast, once temperate phages infect host cells, they can either kill the host cell thereby releasing viral progeny or integrate their genetic material with that of their bacterial host. Once the viral genome of a temperate phage has integrated, the bacterium is referred to as a lysogen. In the lysogenic stage, minimal transcription and translation of viral proteins occurs and the viral genome ("prophage") is transmitted vertically. Later, induction of the prophage can occur and the virus can re-enter the lytic pathway. The choice of whether to lyse a host cell or enter a latent state upon infection as well as the mechanisms of prophage induction have been extensively studied in temperate phages. Analysis of the molecular mechanisms underlying the lysis vs. lysogeny pathways has formed the basis for formative work on gene regulation [146]. The study of the lysis-lysogeny switch in phage λ has become a model system for understanding how temperate phages "decide" the means by which they exploit hosts [101, 74, 8, 11, 97, 137, 20, 40, 49, 98, 131, 153].

Nearly all theoretical studies of temperate phage decision dynamics have claimed

[1]This chapter appeared in the Biophysical Journal [193].

that switching between alternative pathways depends on some change in environmental conditions or some other random process inherent to the virus [8, 11, 159, 20]. However, long-standing experimental assays [101] and recent molecular studies [97, 40] have found that changes in multiplicity of infection drastically change the initial decision switch in phage λ. The experimental evidence indicates that two or more coinfecting phages will lead to lysogeny, whereas a single infecting λ-phage leads to cell lysis [101, 97, 40]. This qualitative change in outcome based on small differences in levels of coinfection may seem counter-intuitive, and thus far, lacks a coherent theoretical description. Here we show that when multiple viruses infect the same host cell, nonlinearities in gene regulatory dynamics can lead to qualitative changes in steady state gene expression, and ultimately to a deterministic outcome, *i.e.* lysis or lysogeny. This finding suggests that features of viral collective decision switches are inheritable, mutable, and evolvable. The evolvable quantities in this process include the critical multiplicity of infection at which the switch occurs and the ratio of steady state gene expression for the two different decision states. These features are determined, in part, by kinetic parameters of binding which are easily affected by mutation.

2.2 Results

2.2.1 Regulatory model of viral decision processes

Consider a mixture of temperate bacterial viruses and bacterial hosts in which cells are co-infected by $\mathcal{M}$ viruses of an identical strain. A simple model of this situation and a possible mechanism for a viral decision process within a bacterial cell is depicted in Figure 1. Two genes in this network, x and y, share a common promoter region. When X-dimers are bound to the promoter, they halt transcription of gene y and enhance that of x. When Y-dimers are bound to the promoter, they halt transcription of both genes x and y. Hence, the viral regulatory network includes both a positive and a

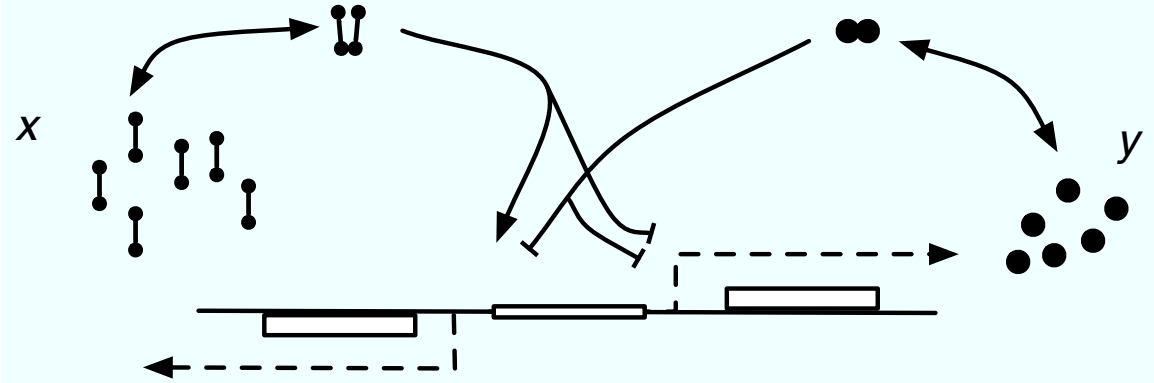

Figure 1: Schematic of a bistable switch in which dimers of an excitatory loop (two circles with line) and an inhibitory loop (circles) compete for the same promoter. Solid lines denote protein interactions and dashed lines denote transcriptional events.

Table 1: Transcription rates of mRNAs given promoter sites that are unoccupied or occupied by either X or Y dimers. The dimer Y is strictly inhibitory whereas X both represses Y and activates itself. When $\alpha_y > \alpha_x$ the inhibitory loop is favored at low $\mathcal{M}$.

Occupancy	x transcription	y transcription
None	α_x	α_y
X dimer	β_x	0
Y dimer	0	0

negative feedback loop (see Table 1 for more details).

Unlike other models in which the copy number remains fixed, here we set the copy number of viral genes equal to the number of coinfecting viruses, $\mathcal{M}$. Following standard methods [119, 33, 71, 20], we consider a mass-action model of the dynamics of promoters, mRNA, and proteins. The network is comprised of an excitatory and an inhibitory loop. When promoters are unoccupied, mRNA is transcribed at a rate α_x and α_y for the excitatory and inhibitory loops respectively. These monomers can dimerize and then bind to the promoter site. When X-dimers are bound, they inhibit transcription of gene y and enhance transcription of gene x at a rate β_x. When Y-dimers are bound, they inhibit transcription of both genes x and y. Following standard methods in the field [20], we track the density of monomers x_1 and y_1,

dimers x_2 and y_2, mRNA transcripts m_x and m_y, and promoter occupancies d_0, d_x, and d_y, such that the dynamics are

$$\begin{aligned}
\dot{x_1} &= 2\kappa_- x_2 - 2\kappa_+ x_1^2 + \sigma m_x - \gamma_p x_1, && (1)\\
\dot{x_2} &= -\kappa_- x_2 + \kappa_+ x_1^2 + k_- d_x - k_+ d_0 x_2, && (2)\\
\dot{y_1} &= 2\kappa_- y_2 - 2\kappa_+ y_1^2 + \sigma m_y - \gamma_p y_1, && (3)\\
\dot{y_2} &= -\kappa_- y_2 + \kappa_+ y_1^2 + k_- d_y - k_+ d_0 y_2, && (4)\\
\dot{d_0} &= k_- d_x + k_- d_y - k_+ d_0 x_2 - k_+ d_0 y_2, && (5)\\
\dot{d_x} &= k_+ d_0 x_2 - k_- d_x, && (6)\\
\dot{d_y} &= k_+ d_0 y_2 - k_- d_y, && (7)\\
\dot{m_x} &= \alpha_x d_0 + \beta_x d_x - \gamma_m m_x, && (8)\\
\dot{m_y} &= \alpha_y d_0 - \gamma_m m_y. && (9)
\end{aligned}$$

In this model, κ_- and κ_+ are the rates of unbinding and dimerization of monomers, k_- and k_+ are the rates of detachment and attachment of dimers to promoter sites, σ is the translational rate, $\alpha_{x,y}$ and β_x are transcriptional rates, γ_m is the degradation rate of transcripts, and γ_p is the degradation rate of proteins. Note that the total concentration of promoter sites remains unchanged, and denote that as $d = \mathcal{M}C$, where $\mathcal{M}$ is the cellular multiplicity of infection and C is a conversion factor corresponding to the molar concentration of a single molecule in a volume equivalent to a bacterial cell. Then $d = d_0 + d_x + d_y$ is a constant throughout the dynamics.

In this complete form, mathematical analysis is impractical. Hence, we apply a non-restrictive quasi-steady state approximation (QSSA) to the full model. (see Section 2.4.1 for a detailed treatment of this derivation). In the QSSA, we take advantage of the disparity in rates between fast and slow processes in the gene regulatory network and assume that other variables are determined by the slowly varying monomer concentrations. We are able to derive expressions for the rate of change of the rescaled

concentrations of X and Y free monomers, denoted here as u and v, respectively,

$$\dot{u} = \mathcal{M}\frac{\alpha_u + \beta_u u^2}{1 + u^2 + v^2} - \gamma_p u, \tag{10}$$

$$\dot{v} = \mathcal{M}\frac{\alpha_v}{1 + u^2 + v^2} - \gamma_p v. \tag{11}$$

The parameters α_u, α_v, and β_u are rescaled rates combining the effects of binding, transcription, translation, and degradation where γ_p is the protein degradation rate (see Section 2.4.1 for definitions). Importantly, the reformulation of the model shows that changes in the cellular multiplicity of infection, $\mathcal{M}$, directly affect the rates of transcription (in the full model of Equations (1–9)) and translation (in the protein-only model of Equations (10–11)). Thus, changes in the number of viruses within a cell affect the values of key parameters in a nonlinear dynamic model of regulatory control. Further, predictions of steady state expression are equivalent whether we are considering the full dynamics of promoters, mRNA and proteins or the dynamics of proteins in Equations (10–11).

The steady state monomer concentrations in this model, $(\bar{u}, \bar{v})$, can be solved implicitly. After some analysis (see Section 2.4.1), we find exact expressions for $\mathcal{M}(\bar{u})$ and $\bar{v}(\bar{u})$,

$$\mathcal{M} = \gamma_p \bar{u}\left(\frac{1 + \bar{u}^2}{\alpha_u + \beta_u \bar{u}^2} + \frac{\alpha_v^2 \bar{u}^2}{(\alpha_u + \beta_u \bar{u}^2)^3}\right), \tag{12}$$

$$\bar{v} = \frac{\alpha_v \bar{u}}{\alpha_u + \beta_u \bar{u}^2}. \tag{13}$$

The ratio, $\bar{v}/\bar{u}$, shows the disparity in steady state expressions and is equivalent to the ratio of the actual concentrations. For low values of $\bar{u}$, $\bar{v}/\bar{u}$ is on the order of α_v/α_u, whereas for large values of $\bar{u}$, $\bar{v}/\bar{u}$ is on the order of $\alpha_v/(\beta_u \bar{u}^2)$. When baseline production of the inhibitory protein exceeds that of the excitatory protein ($\alpha_v > \alpha_u$), the ratio of steady state expression switches from high to low as $\bar{u}$ increases. This is the key to the origin of collective decision making in viruses. When $\mathcal{M}$ is low, total expression is kept at low levels and the decision switch favors the inhibitory

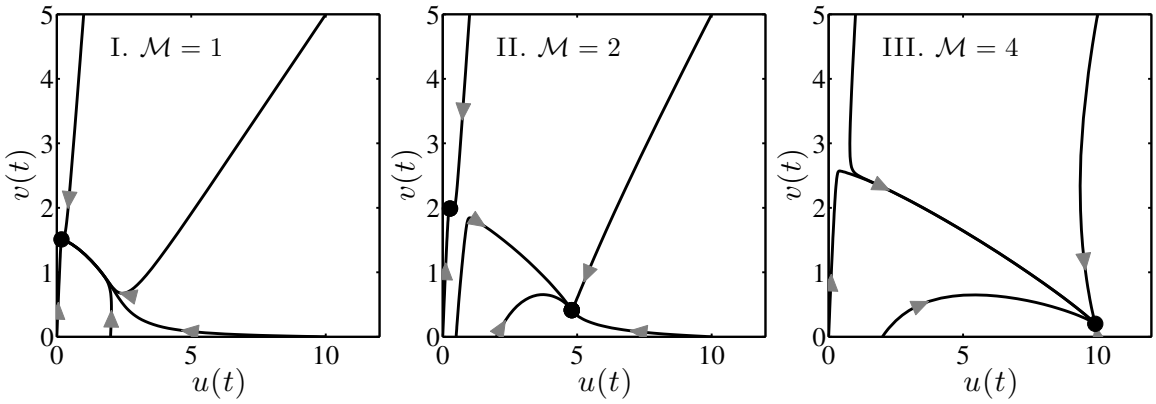

Figure 2: Phase plane dynamics of $v(t)$ vs. $u(t)$ in the protein only model of Equations (10–11) given $\alpha_u = 0.5$, $\alpha_v = 5$, $\beta_u = 2.5$, $\gamma_p = 1$, and $\mathcal{M} = 1$, 2, and 4 respectively. Arrows denote direction of trajectories and solid circles are stable equilibria. From left to right, figures depict the three regimes of the model where dynamics are dominated by the inhibitory pathway (v), contains alternative steady states, and are dominated by the excitatory pathway (u), respectively.

loop. When $\mathcal{M}$ is large, total expression increases and the decision switch favors the excitatory loop for which the nonlinear positive feedback dominates.

There can be at most three outcomes for this model: (i) an inhibition-dominated regime; (ii) a bistable regime; (iii) an excitation-dominated regime (see Figure 2). These regimes occur for $\mathcal{M} < \mathcal{M}_1$, $\mathcal{M}_1 < \mathcal{M} < \mathcal{M}_2$ and $\mathcal{M} > \mathcal{M}_2$ respectively (see Figure 3). The critical values, $\mathcal{M}_1$ and $\mathcal{M}_2$, indicate the cellular multiplicity of infection when a change in behavior is expected. We cannot find explicit solutions for $\mathcal{M}_{1,2}$ as a function of the parameters. However, implicit solutions are possible since these critical points satisfy the condition $\partial\mathcal{M}(\bar{u})/\partial\bar{u} = 0$ in Equation (12). Analysis demonstrates that when $\beta_u \gg \alpha_u$ there must be two critical points where saddle node bifurcations occur. First, when $\mathcal{M} = \mathcal{M}_1$, an unstable and stable state emerge. Next, when $\mathcal{M} = \mathcal{M}_2$, the unstable equilibrium collides with the other stable branch. When $\alpha_v > \alpha_u$, there is only one possible steady state for low cellular multiplicity of infection ($\mathcal{M} < \mathcal{M}_1$), where $\bar{v} \gg \bar{u}$. Likewise, there is only one possible steady state for high cellular multiplicity of infection ($\mathcal{M} > \mathcal{M}_2$), where $\bar{u} \gg \bar{v}$. Thus, a cell's fate

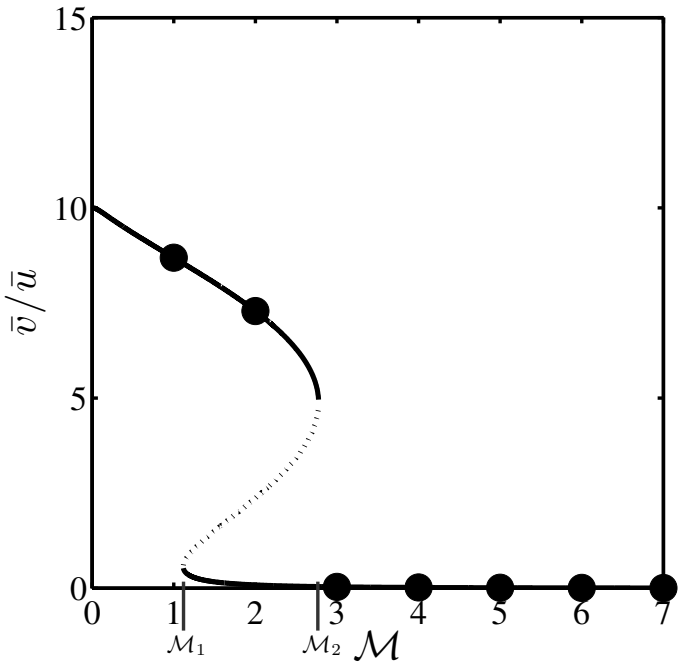

Figure 3: Ratio of steady state protein concentrations for varying cellular multiplicity of infection ($\mathcal{M}$) in the protein only model of Equations (10–11) given $\alpha_u = 0.5$, $\alpha_v = 5$, $\beta_u = 2.5$ and $\gamma_p = 1$. Solid lines denote stable branches, dotted line denotes unstable branch, and circles are results of numerical simulations given the initial conditions $(u = 0, v = 0)$. There are three regimes in this model $\mathcal{M} < \mathcal{M}_1$, $\mathcal{M}_1 < \mathcal{M} < \mathcal{M}_2$, and $\mathcal{M} > \mathcal{M}_2$ where the model is dominated by the inhibitory pathway (v), contains alternative steady states, and is dominated by the excitatory pathway (u), respectively. The critical points $\mathcal{M}_1$ and $\mathcal{M}_2$ correspond to saddle-node bifurcations of the model.

can be deterministically tuned by, in some cases, the addition or subtraction of a few infecting viruses. In the intermediate regime, $\mathcal{M}_1 < \mathcal{M} < \mathcal{M}_2$, the outcome depends strongly on stochastic effects which may drive the system to one stable expression state or the other. Alternatively, if we had tracked the total concentration of proteins, equal to the sum of monomers, dimers, and bound dimers, we would find the same critical values of $\mathcal{M}$ for bifurcations.

The central elements of this model are the transcriptional feedback and protein dimerization, as has been pointed out in other contexts [33]. Without feedback, increases in copy number would not qualitatively change the ratio of gene expression.

Without dimerization, the ratio of gene expression would change with varying $\mathcal{M}$, but the change would not be as drastic and there would no longer be a sequence of bifurcations. Given feedback and dimerization, the finding of a sequence of copy-number driven bifurcations are robust to changes in the kinetic parameters. Changes in kinetic parameters will alter features of the copy-number controlled bifurcation, including the critical values of $\mathcal{M}$ where the bifurcations occur and the relative change in expression before and after the bistable regime.

2.2.2 Mechanistic model of the phage λ initial decision switch

The generic mechanism presented for a deterministically controlled decision switch also applies in more complex scenarios. A simplified version of the lysis-lysogeny switch in phage λ involving genes *cI*, *cro*, and *cII* is presented in Figure 4. It is now widely believed that the ultimate fate of the decision process in phage λ is controlled by CII. High levels of CII facilitate transcription of CI (and the lysogenic pathway) whereas low levels of CII favor production of Cro (and the lytic pathway) [146, 97, 40, 98]. The molecular mechanism proposed is that the amount of CII proteins may indirectly "measure" the number of infecting viruses [97]. Despite extensive experimental evidence that increases in coinfection systematically switches a cell's fate from lysis to lysogeny, as yet there is no general theory that explains this process. In particular, why do increases in the copy numbers of viral genes not lead to a proportional increase in the concentration of all components of the regulatory system in a way that their ratios (and hence decisions) are left unchanged?

In the model proposed here, we do not consider the entire phage λ decision circuit. Rather, we propose a simplified model that captures critical features of the decision switch controlled at multiple promoter sites [146]. The two promoter sites considered here are denoted as (i) P_{RM}/P_R and (ii) P_{RE}, see Figure 4 for more details. In addition, CI, Cro and CII all dimerize prior to binding. The rules of transcription

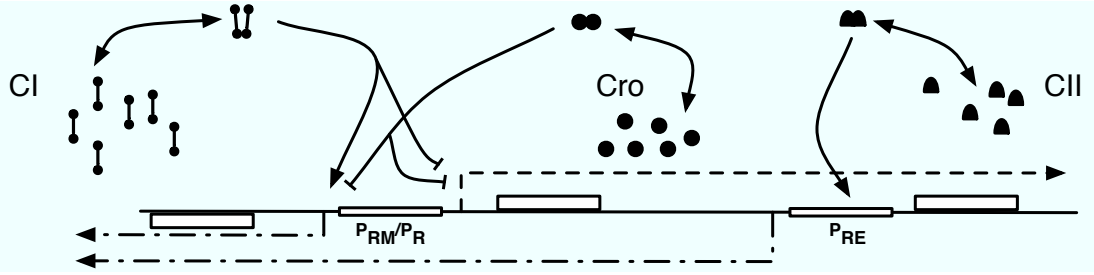

Figure 4: Simplified version of the phage λ switch in which CII acts as a gate protein between the CI and the Cro pathways [146, 40]. In the schematic, solid lines denote protein interactions, the dashed line denotes transcription events that require no activation, and dashed-solid lines denote transcription events that require activation. More details can be found in Section 2.4.2.

are as follows: (i) baseline transcription initiates at P_{RM}/P_R to make *cro* and *cII* transcripts; (ii) binding of CII at P_{RE} leads to transcription of *cI*; (iii) binding of CI at P_{RM}/P_R catalyzes transcription of *cI* and inhibits transcription of *cro* and *cII*; and (iv) binding of Cro at P_{RM}/P_R inhibits transcription of all genes. In reality, the P_{RM} and P_R promoters are distinct and comprised of an overlapping set of three operator regions, which we ignore in the interest of analytic tractability [2, 163, 146].

We model this system, as before, by tracking the dynamics of promoter, mRNA and protein concentrations. A full list of dynamical expressions and parameter values can be found in Section 2.4.2. As in the generic model, we derive equations approximating the dynamics of protein concentration. Denoting u, v, and w as the rescaled concentrations of CI, Cro and CII monomers we find the rates of change to be:

$$\dot{u} = \mathcal{M}\frac{\beta_u u^2}{1+u^2+v^2} + \mathcal{M}\frac{\delta_u w^2}{1+w^2} - \gamma_u u, \tag{14}$$

$$\dot{v} = \mathcal{M}\frac{\alpha_v}{1+u^2+v^2} - \gamma_v v, \tag{15}$$

$$\dot{w} = \mathcal{M}\frac{\alpha_w}{1+u^2+v^2} - \gamma_w w, \tag{16}$$

where $\mathcal{M}$ is, as before, the cellular multiplicity of infection and the rescaled variables are defined in Section 2.4.2.

The values of these rescaled parameters depend on kinetic rates, some of which

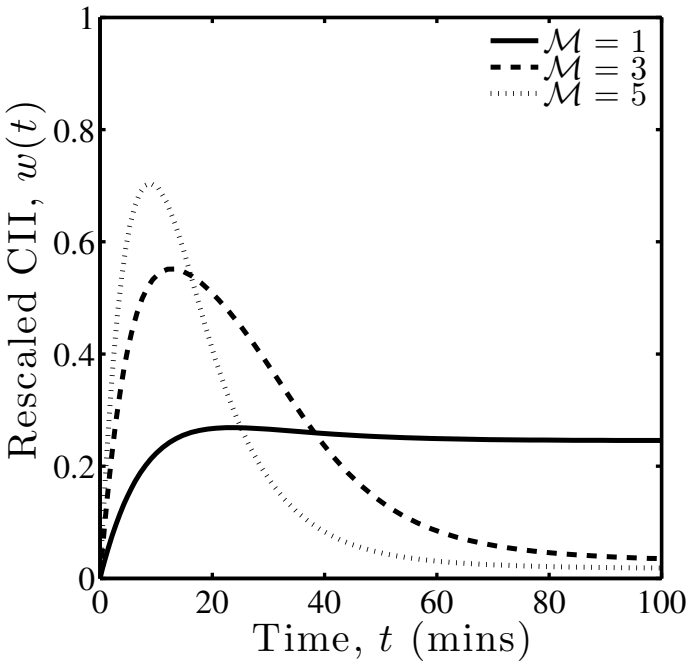

Figure 5: Simulated dynamics of the decision switch as a function of multiplicity of infection, where $\mathcal{M} = 1, 3$ and 5 correspond to solid, dashed, and dotted lines respectively. Notice that CII functions as a gating protein. Increases in $\mathcal{M}$ shift CII above a critical threshold enabling transcription of CI and coupling to the nonlinear positive feedback loop. Values of rescaled parameters used in the numerical simulations are $\beta_u = 0.08$, $\delta_u = 0.06$, $\alpha_v = 0.04$, $\alpha_w = 0.04$, $\gamma_u = 0.04$, $\gamma_v = 0.05$, and $\gamma_w = 0.12$ all in units of min^{-1}. More details can be found in Section 2.4.2.

have been studied in the literature while others have not. Generically, the CI-CII-Cro model undergoes a series of saddle-node bifurcations that go from a stable regime where Cro dominates ($\bar{v} > \bar{u}$) to a bistable regime and back to a stable regime where CI dominates ($\bar{u} > \bar{v}$). CII acts as a gate protein in this system. Increasing $\mathcal{M}$ drives the dynamic level of w past a critical point where the nonlinearity of the positive feedback in *CI* production leads to a switch in behavior (see Figures 5–6). These results are robust to small changes in the values of parameters, and so many of the values in the regulatory network could be changed and the switch would still function. As in the previous case, the system possesses only one steady state given suitably low or high levels of coinfection. Hence, the switch is dominated by deterministic behavior

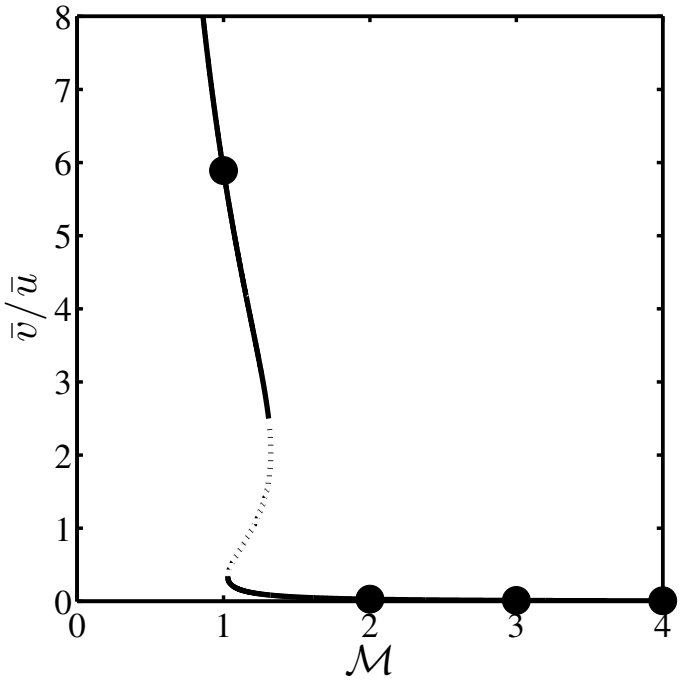

Figure 6: Bifurcation diagram for $\bar{v}/\bar{u}$ as a function of $\mathcal{M}$ where solid line is analytical curve and circles are from numerical simulations of Equations (14–16) given the initial conditions, $(u = 0, v = 0, w = 0)$**.** Values of rescaled parameters are the same as those used in Figure 5.

in contrast to suggestions that decision outcomes must have stochastic origins or be driven by changes in environmental conditions [8].

Parameters are chosen (see Section 2.4.2) such that maximal transcriptional rates are on the order of a few transcripts per minute per gene, maximal translation rates are on the order of one protein for every few minutes per transcript, dissociation constants are on the order of 10^7 M^{-1}, and mRNA and protein degradation rates are on the order of 0.1 per minute [2, 8, 11, 159]. Figures 5–6 demonstrate that these parameter values can lead to a $\mathcal{M}$-driven switch where lysis is favored at $\mathcal{M} = 1$ and lysogeny is favored at $\mathcal{M} \geq 2$ in agreement with observations [101, 97, 146]. Peak concentrations of CI and Cro are of the same magnitude as experimental observations (in the hundreds of molecules) [146]. Quantitative deviations are unsurprising given the uncertainty in parameters and the reduction of network complexity as compared

to the hypothesized phage λ switch. The relationship between kinetic parameters and features of a genetic switch controlled by copy number, including the width of the bistable regime and the difference in the steady state expression of the two bistable states, are explored in Chapter 3. Note also in this model, if the degradation rate of CI is increased suddenly, as occurs after a cell experiences DNA damage, then the lysogenic state becomes unstable in agreement with experiments [146] and other numerical studies [11, 159, 184, 20].

2.3 Discussion

Living organisms exhibit a remarkable range of complex group behaviors [27, 125, 195]. Often, the number of individuals is a key factor in determining when and if the group exhibits functions and/or behaviors distinct from those of individuals [186, 125, 41, 195]. Although viruses are one of the few types of organisms for which collective decision making has not yet been proved, we have demonstrated a mechanism by which they may do so. The finding that multiple infections can change behavior within a cell indicates that viral infections are not static, but rather may react to their own dynamics. This response is, in principle, an evolvable life history trait of bacterial viruses conferring some selective benefit to strains which adopt this strategy. The ecological circumstances favoring lysogeny have been addressed previously, though the issue is far from settled [173]. Although phage λ often constitutes the dogma for how a temperate phage behaves, mutations that lead to changes in network structure, degree of cooperativity, and kinetic rates could lead to qualitative shifts in the exploitation strategy of a host [97]. Viruses could kill at low $\mathcal{M}$ and go latent at high $\mathcal{M}$, or vice-versa, depending on binding parameters in the regulatory feedback loops controlling a cell's fate. Virulent bacteriophages differ in their life history traits (such as burst size or virion decay rate) by multiple orders of magnitude [45], which suggests that diverse life history strategies may be found in temperate bacteriophages as well. Indeed,

observations of phage λ mutants have already shown that decision variants may in fact be engineered [109, 10].

There are many challenges remaining in the study of multiplicity of infection in bacterial viruses. First, the dynamics of intracellular mRNAs and proteins are stochastic, and we will use a stochastic version of the lysis-lysogeny decision switch evaluate the likelihood of chance outcomes in a deterministically-driven decision system in Chapter 4. Next, the infection of a host cell is rarely simultaneous. Therefore, the subsequent infection by viruses leads to discrete shifts in dynamics and in the parameters controlling the unfolding of exploitation. In that sense, the ecological dynamics of infection and the intracellular dynamics of decision making are necessarily coupled. To what extent subsequent viruses can change the outcome of a previously infected, but not yet committed, virus remains an open question. Finally, the decision-switch we presented is a simplification of many complex intracellular processes. Analyses of the decision switch that incorporate additional biological realism should retain the feature of sensitivity to $\mathcal{M}$. We expect that our finding of nonlinear copy number effects will remain a necessary part of subsequent models.

Experimental tests of the effect of multiplicity of infection on the lysis-lysogeny switch have been conducted using plate-based assays at the population level [101] and using expression fluorescence assays of synchronously infected cell ensembles [97]. We believe that single-cell experiments are necessary in order to test the nonlinear effect of copy number on decision outcome [14, 5]. Simultaneous measurements of viral coinfection level and expression dynamics will facilitate unambiguous determination of the link between $\mathcal{M}$ and the genetic cascade leading to lysis or lysogeny. Already, experimental monitoring of single-cell expression dynamics has provided insight into the sequence of events that allow the λ prophage to induce upon UV irradiation [5]. Copy number variation may affect the initial decision switch as well as prophage induction. For example, variation in physiological state will lead to dynamic changes

in the number of chromosomes within a lysogen. Such dynamic changes could modify rates of viral gene expression affecting prophage function in the presence and absence of cell stress [153].

The choice of whether to burst from a cell or to remain/enter a latent state is a key feature of viruses from phages to human pathogens. The cellular multiplicity of infection may well play a role in shaping other viral decision processes, even if the genetic details are different. At the very least, we have shown here that small changes in coinfection are a sufficient determinant of the initial lysis vs. lysogeny switch upon infection. More generally, our findings suggest that other gene regulatory modules may depend sensitively on copy number, by modifying kinetic rates of transcription within a nonlinear dynamical system.

2.4 Methods and Models

2.4.1 Bistable switch dynamics using a quasi-steady state approximation

Consider a model of viral exploitation in which two competing regulatory pathways share a common promoter. The network is comprised of an excitatory and an inhibitory loop as described in the main text. We assume that translation and transcription are the slow processes in the model and binding and dimerization are fast. Hence, variables x_2, y_2, d_x, d_y, and d_0 are changing much faster than the rest. Therefore, we obtain a quasi-steady state approximation (QSSA) of the full system by setting the corresponding derivatives to zero:

$$0 = -\kappa_- x_2 + \kappa_+ x_1^2 + k_- d_x - k_+ d_0 x_2, \tag{17}$$

$$0 = -\kappa_- y_2 + \kappa_+ y_1^2 + k_- d_y - k_+ d_0 y_2, \tag{18}$$

$$0 = k_+ d_0 x_2 - k_- d_x, \tag{19}$$

$$0 = k_+ d_0 y_2 - k_- d_y. \tag{20}$$

Since we are interested in steady states in this model we omit a more careful treatment of pre-factors involved with this QSSA as presented elsewhere [20]. Adding

Equation (17) and Equation (19), Equation (18) and Equation (20) we obtain

$$0 = -\kappa_- x_2 + \kappa_+ x_1^2, \tag{21}$$

$$0 = -\kappa_- y_2 + \kappa_+ y_1^2, \tag{22}$$

which implies that

$$x_2 = c_p x_1^2, \tag{23}$$

$$y_2 = c_p y_1^2, \tag{24}$$

where $c_p = \kappa_+/\kappa_-$.

Recall that the total concentration of promoter sites remains unchanged, hence $d = d_0 + d_x + d_y$ is a constant throughout the dynamics. After some algebra we find that

$$d_x = d\frac{c_p c_d x_1^2}{1 + c_p c_d (x_1^2 + y_1^2)}, \tag{25}$$

$$d_y = d\frac{c_p c_d y_1^2}{1 + c_p c_d (x_1^2 + y_1^2)}, \tag{26}$$

$$d_0 = \frac{d}{1 + c_p c_d (x_1^2 + y_1^2)}, \tag{27}$$

where $c_d = k_+/k_-$.

Thus, the quasi-steady state approximation for the system has the following form:

$$\dot{x_1} = \sigma m_x - \gamma_p x_1, \tag{28}$$

$$\dot{y_1} = \sigma m_y - \gamma_p y_1, \tag{29}$$

$$\dot{m_x} = d\frac{\alpha_x + \beta_x c_p c_d x_1^2}{1 + c_p c_d (x_1^2 + y_1^2)} - \gamma_m m_x, \tag{30}$$

$$\dot{m_y} = d\frac{\alpha_y}{1 + c_p c_d (x_1^2 + y_1^2)} - \gamma_m m_y. \tag{31}$$

We simplify the system further by assuming transcription is faster than translation. Setting the derivatives of m_x and m_y to zero yields

$$m_x = \frac{d}{\gamma_m}\frac{\alpha_x + \beta_x c_p c_d x_1^2}{1 + c_p c_d (x_1^2 + y_1^2)}, \tag{32}$$

$$m_y = \frac{d}{\gamma_m}\frac{\alpha_y}{1 + c_p c_d (x_1^2 + y_1^2)}. \tag{33}$$

Substituting these expressions into the equations for x_1 and y_1 we get the following system:

$$\dot{x}_1 = \frac{d\sigma}{\gamma_m}\frac{\alpha_x + \beta_x c_p c_d x_1^2}{1 + c_p c_d (x_1^2 + y_1^2)} - \gamma_p x_1, \tag{34}$$

$$\dot{y}_1 = \frac{d\sigma}{\gamma_m}\frac{\alpha_y}{1 + c_p c_d (x_1^2 + y_1^2)} - \gamma_p y_1. \tag{35}$$

We then rescale the variables: $u = \sqrt{c_p c_d}x_1$, $v = \sqrt{c_p c_d}y_1$, and set $\beta_u = C\beta_x\sigma\sqrt{c_p c_d}/\gamma_m$, $\alpha_u = C\alpha_x\sigma\sqrt{c_p c_d}/\gamma_m$, and $\alpha_v = C\alpha_y\sigma\sqrt{c_p c_d}/\gamma_m$. The dynamics of the rescaled expression are

$$\dot{u} = \mathcal{M}\frac{\alpha_u + \beta_u u^2}{1 + u^2 + v^2} - \gamma_p u, \tag{36}$$

$$\dot{v} = \mathcal{M}\frac{\alpha_v}{1 + u^2 + v^2} - \gamma_p v. \tag{37}$$

where, as noted before, $\mathcal{M} = d/C$ is the copy number of viral genes. This system can be thought of as a protein only model of the switch.

Continuing our derivation of steady states, we set the above derivatives to zero:

$$0 = \mathcal{M}\frac{\alpha_u + \beta_u u^2}{1 + u^2 + v^2} - \gamma_p u, \tag{38}$$

$$0 = \mathcal{M}\frac{\alpha_v}{1 + u^2 + v^2} - \gamma_p v. \tag{39}$$

The second equation implies that

$$\frac{\mathcal{M}}{1 + u^2 + v^2} = \frac{\gamma_p}{\alpha_v}v. \tag{40}$$

Substituting this expression into the first equation we get

$$\frac{\gamma_p}{\alpha_v}v(\alpha_u + \beta_u u^2) = \gamma_p u \implies \bar{v} = \frac{\alpha_v \bar{u}}{\alpha_u + \beta_u \bar{u}^2}. \tag{41}$$

From this, we can now find the expression for $\mathcal{M}$ as a function of $\bar{u}$:

$$\mathcal{M} = \gamma_p \bar{u}\left(\frac{1 + \bar{u}^2}{\alpha_u + \beta_u \bar{u}^2} + \frac{\alpha_v^2 \bar{u}^2}{(\alpha_u + \beta_u \bar{u}^2)^3}\right). \tag{42}$$

This function defines implicit dependence of $\bar{u}$ on $\mathcal{M}$. We can find values of $\mathcal{M}$ at which $\partial\mathcal{M}(\bar{u})/\partial\bar{u} = 0$, which correspond to critical values where qualitative changes

in steady state expression are expected. We only consider monomer degradation in this and the subsequent model in Section 2.4.2. Assuming only monomers degrade, only dimers degrade, or some combination of those scenarios does not change any of the qualitative features of our results.

2.4.2 Dynamics of CI-CII-Cro

The dynamics of the CI-CII-Cro system as depicted in Figure 4 of the main text are:

$$\begin{aligned}
\dot{x_1} &= 2\kappa_-^{(x)} x_2 - 2\kappa_+^{(x)} x_1^2 + \sigma m_x - \gamma_x x_1, && (43)\\
\dot{x_2} &= -\kappa_-^{(x)} x_2 + \kappa_+^{(x)} x_1^2 + k_-^{(x)} d_x - k_+^{(x)} d_0 x_2, && (44)\\
\dot{y_1} &= 2\kappa_-^{(y)} y_2 - 2\kappa_+^{(y)} y_1^2 + \sigma m_y - \gamma_y y_1, && (45)\\
\dot{y_2} &= -\kappa_-^{(y)} y_2 + \kappa_+^{(y)} y_1^2 + k_-^{(y)} d_y - k_+^{(y)} d_0 y_2, && (46)\\
\dot{z_1} &= 2\kappa_-^{(z)} z_2 - 2\kappa_+^{(z)} z_1^2 + \sigma m_z - \gamma_z z_1, && (47)\\
\dot{z_2} &= -\kappa_-^{(z)} z_2 + \kappa_+^{(z)} z_1^2 + k_-^{(z)} e_z - k_+^{(z)} e_0 z_2, && (48)\\
\dot{d_0} &= k_-^{(x)} d_x + k_-^{(y)} d_y - k_+^{(x)} d_0 x_2 - k_+^{(y)} d_0 y_2, && (49)\\
\dot{d_x} &= k_+^{(x)} d_0 x_2 - k_-^{(x)} d_x, && (50)\\
\dot{d_y} &= k_+^{(y)} d_0 y_2 - k_-^{(y)} d_y, && (51)\\
\dot{e_z} &= k_+^{(z)} e_0 z_2 - k_-^{(z)} e_z, && (52)\\
\dot{e_0} &= -k_+^{(z)} e_0 z_2 + k_-^{(z)} e_z, && (53)\\
\dot{m_x} &= \beta_x d_x + \delta_x e_z - \gamma_m m_x, && (54)\\
\dot{m_y} &= \alpha_y d_0 - \gamma_m m_y, && (55)\\
\dot{m_z} &= \alpha_z d_0 - \gamma_m m_z. && (56)
\end{aligned}$$

Using a sequence of quasi-steady state approximations in which we integrate over processes of dimerization, binding and mRNA production, the model can be reduced to a protein only dynamic model of x, y, and z monomer concentration. Here we relax the assumption that dimerization and binding constants are identical for each

protein. In analogy to the two protein system derived earlier the model equations are

$$\begin{aligned}
\dot{x} &= \frac{\mathcal{M}C\beta_x\sigma/\gamma_m\left(c_p^{(x)}c_d^{(x)}x^2\right)}{1+c_p^{(x)}c_d^{(x)}x^2+c_p^{(y)}c_d^{(y)}y^2}\cdots \\
&\quad +\frac{\mathcal{M}C\delta_x c_p^{(z)}c_d^{(z)}z^2}{1+c_p^{(z)}c_d^{(z)}z^2}-\gamma_x x, \qquad (57)\\
\dot{y} &= \frac{\mathcal{M}C\alpha_y\sigma/\gamma_m}{1+c_p^{(x)}c_d^{(x)}x^2+c_p^{(y)}c_d^{(y)}y^2}-\gamma_y y, \qquad (58)\\
\dot{z} &= \frac{\mathcal{M}C\alpha_z\sigma/\gamma_m}{1+c_p^{(x)}c_d^{(x)}x^2+c_p^{(y)}c_d^{(y)}y^2}-\gamma_z z, \qquad (59)
\end{aligned}$$

where $c_p^{(x)}=\kappa_+^{(x)}/\kappa_-^{(x)}$, $c_d^{(x)}=k_+^{(x)}/k_-^{(x)}$ and likewise for y and z. In addition, here $\mathcal{M}$ denotes the integer multiplicity of infection and C is the conversion factor for representing operator concentrations in molar units. As before, the number of promoter sites are given as

$$d_0+d_x+d_y=e_0+e_z=\mathcal{M}C. \qquad (60)$$

Note that converting from concentrations to estimates of molecules per cell has been calibrated based on an assumption of cell volumes on the order of $1-2\ \times 10^{-15}$ L. Hence, the number of molecules in a bacterial cell is equal to the molar concentration divided by C. The dynamic factors involved in tracking the total protein concentration have been studied for the quasi steady state approximation in other contexts [20]. Importantly, they do not alter the predictions regarding steady state behavior.

We propose the following rescaling of this model, $u=\sqrt{c_p^{(x)}c_d^{(x)}}x_1$, $v=\sqrt{c_p^{(y)}c_d^{(y)}}y_1$, $w=\sqrt{c_p^{(z)}c_d^{(z)}}z_1$, $\beta_u=C\beta_x\sigma\sqrt{c_p^{(x)}c_d^{(x)}}/\gamma_m$, $\delta_u=C\delta_x\sigma\sqrt{c_p^{(x)}c_d^{(x)}}/\gamma_m$, $\alpha_v=C\alpha_y\sigma\sqrt{c_p^{(y)}c_d^{(y)}}/\gamma_m$, $\alpha_w=C\alpha_z\sigma\sqrt{c_p^{(z)}c_d^{(z)}}/\gamma_m$, $\gamma_u=\gamma_x$, $\gamma_v=\gamma_y$, and $\gamma_w=\gamma_z$. Using this rescaling we recover the model in Equations (14–16) in the main text. Parameter estimates are approximate. They are in range with experimental measurements and typical values for dimerization, binding, transcription, translation and degradation in bacteria and viruses. The approximate kinetic values are as follows [2, 11, 159]: $c_p^{(x)}\approx 10^7\mathrm{M}^{-1}$, $c_p^{(y)}\approx 10^7\mathrm{M}^{-1}$, $c_p^{(z)}\approx 10^7\mathrm{M}^{-1}$, $c_d^{(x)}\approx 10^7\mathrm{M}^{-1}$, $c_d^{(y)}\approx 10^7\mathrm{M}^{-1}$, $c_d^{(z)}\approx 10^7\mathrm{M}^{-1}$,

$\beta_x \approx 1.6\text{min}^{-1}$, $\delta_x \approx 1.2\text{min}^{-1}$, $\alpha_y \approx 0.8\text{min}^{-1}$, $\alpha_z \approx 0.8\text{min}^{-1}$, $\sigma \approx 0.5\text{min}^{-1}$, $\gamma_m \approx 0.1\text{min}^{-1}$, $\gamma_x \approx 0.04\text{min}^{-1}$, $\gamma_y \approx 0.05\text{min}^{-1}$, $\gamma_z \approx 0.12\text{min}^{-1}$, and $C \approx 10^{-9}$ M. Here the degradation of CII is higher than either CI or Cro, in part due to the instability of the protein and the need for other components such as CIII that prevent its degradation by bacterial proteases [146]. Importantly, model behaviors are robust to small changes in these parameters, and, in particular, to changes in the rescaled values of β_u, δ_u, α_v, α_w and the degradation rates $\gamma_{u,v,w}$.

CHAPTER III

SMALL-SCALE COPY NUMBER VARIATION AND LARGE-SCALE CHANGES IN GENE EXPRESSION[1]

3.1 Introduction

Copy number variation (CNV) is an important and widespread component of within and between population genetic variation. The copy number of genes and gene fragments varies significantly over physiological to evolutionary time scales with multiple effects on phenotype. For example, CNV can cause statistically significant changes in concentrations of RNA associated with growth rate changes in bacteria [95, 172] and in enzyme concentrations associated with nutrient intake in humans [142, 47]. The copy number of viral genomes undergoes dynamical changes during multiple infection of bacteria by phages, leading to qualitative changes in gene regulation which may lead to alternative modes of exploitation [97, 193]. The duplication of a gene can facilitate subsequent diversification - a mechanism considered to be a dominant cause of phenotypic innovation [136, 111, 107, 100]. In extreme cases, whole-genome duplications have led to lineage diversification within yeast [92]. In humans, large-scale deletions and duplications of chromosomes are known to cause severe genetic disorders [154, 37] and are imputed in the onset of other diseases including cancer [143, 144]. Finally, multiple studies have demonstrated that CNV in humans is far more extensive than previously believed, though its impact on phenotype is yet to be fully resolved [149, 17, 38, 175, 93].

Despite its ubiquity, CNV has been nearly universally overlooked in quantitative

[1]This chapter appeared in the Proceedings of the National Academy of Sciences of the United States of America [124].

models of gene regulation. In those cases where quantitative models of CNV have been developed, the primary focus has been on changes in the copy number itself, as in the case of plasmid maintenance [140] and dynamics of transposable elements [24]. In some instances, gene copy number is integrated into dynamic models of regulation to account for cell-to-cell variability of regulatory elements found on plasmids [68]. More commonly, recent studies have attempted to identify statistical relations between CNV and fitness [47], protein interactions [145], or combinations of both [66]. In order to understand the progression from CNV to changes in phenotype to changes in fitness, it seems necessary to carefully examine gene regulation itself. The dynamics of a gene regulatory network depends on network topology, the quantitative nature of feedbacks and interactions between DNA, RNA and proteins, epigenetic modifications of regulatory elements, and the biochemical state of the intracellular and surrounding environment. Additionally, as we argue here, gene regulatory dynamics can also depend sensitively on the copy number of genes and promoters. For example, in synthetically designed networks, small changes in the copy number of gene regulatory modules have been shown to lead to qualitative changes in gene expression [9, 82]. In naturally occurring networks, there may be selection pressures on kinetic parameters such that normally occurring levels of copy number are far from or close to the critical threshold that would lead to a dramatic change in gene expression.

In this chapter, we take a quantitative approach to assess when small changes in copy number can have a dramatic, nonlinear effect on gene expression. We study the effect of changing copy number within a series of small, regulatory networks commonly referred to as "network motifs" [126]. These motifs are network subgraphs shown to be building blocks of complex regulatory networks [4]. Increasing the number of motifs means that multiple networks are coupled together via a common pool of transcription factors. Changes in the number of promoter sites is directly linked to

changes in the rate of regulated recruitment, which in turn leads to changes in translation and other transcriptional feedbacks [146]. As we demonstrate, small changes in gene copy number within motifs exhibiting positive and/or negative feedback can switch the network from one alternative steady state to another as well as switch gene expression to and from an oscillatory state. Thus, changes in copy number may act as knobs within a nonlinear dynamical system, in much the same way that changes in environmental conditions can drive expression from one steady state to another [71, 4].

3.2 Results

3.2.1 CNV and Network Motifs

We systematically analyze the dependency of four network motifs, positive feedback, bistable feedback, toggle switch, and the repressilator, on the copy number, $\mathcal{N}$. The method for analyzing each of these motifs is largely the same, and illustrated in Figure 7. Although $\mathcal{N}$ is not explicitly present in the mathematical models presented in the second column of Figure 7, it factors in implicitly. Note that $\mathcal{N}$ is proportional to the total concentration of promoter sites, d, i.e. $\mathcal{N} = d/C$, where $C \approx 10^{-9}$M is the conversion factor denoting the molar concentration of a single molecule in the volume of an *E. coli* cell [4]. Hence, from the outset it is evident that copy number can directly modify basic kinetic rates of transcription, binding and unbinding and indirectly modify others as well. In fact, in the quasi-steady state approximation (QSSA) version of all models (see Section 3.4, and Appendix A), the re-scaled translation rates are proportional to copy number. Likewise, changes in cell volume may also have global effects in changing gene expression. The estimate of other kinetic parameters are approximate. They are in range with experimental measurements and typical values for dimerization, binding, transcription, translation and degradation in bacteria and viruses [4] (a list of all parameters used in numerical simulations can be found in

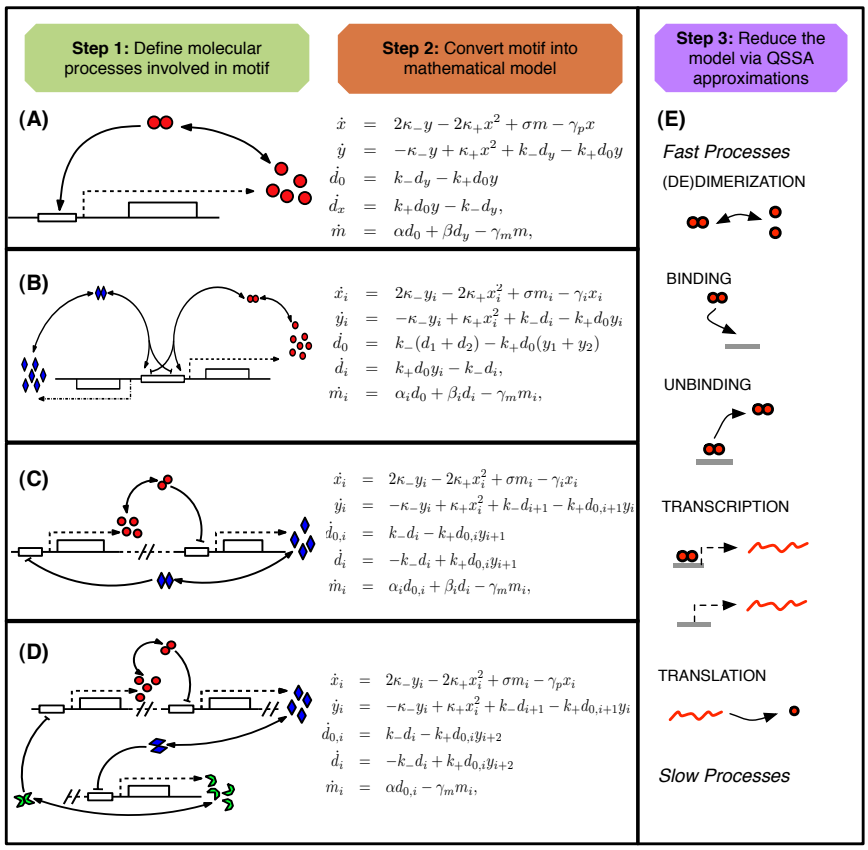

Figure 7: Schematic of the quantitative approach to linking CNV with gene expression in the case of four motifs: (A) Positive feedback; (B) Bistable feedback; (C) Toggle switch; (D) Repressilator. Panel (E) catalogs the process we consider in order from fast to slow processes (degradation of mRNA and proteins are not shown). Variables refer to concentrations of protein monomers (x), protein dimers (y), unoccupied and occupied promoters (d_0 and d_1), and mRNA (m). Parameters $\kappa_\pm$ are the dimerization and de-dimerization rates, $k_\pm$ are the binding and dissociation rates of the dimers to the promoter site, α is the basal transcription rate, β is the regulated transcription rate, σ is the translation rate, γ_m is the degradation rate of mRNA, γ_p and γ_i denote the degradation rate of proteins. The subscript i denotes the index of the promoter, monomer, or dimer in the motif. In (C), the notation for the toggle switch is $i = 1, 2$, and to simplify the index notation we assume that $2+1 = 1$, that is the index wraps back to 1 once it becomes greater than 2. The variables and parameters have the same meaning as in the other cases except that $d_{0,i}$ and d_i denote free and occupied promoters of the i-th gene. For the repressilator system in (D), we again employ our "wrapping" notation, this time using the rule $3 + 1 = 1$, so that when the index becomes greater than 3 it wraps back to 1. Note that in the motif schematics in the first column, transcription and translation are denoted using a single dashed line.

the Appendix A). Note that we do not include degradation of dimers for the sake of analytical tractability, however, numerical tests including degradation of dimers do not qualitatively change any of the results. In presenting results, we emphasize how the steady-state gene expression changes as a function of $\mathcal{N}$. In so doing, we use the term bifurcation to mean a qualitative change in steady-state gene expression [176, 71]. More information on the derivation of the QSSA and bifurcation conditions may be found in Appendix A.

3.2.2 Positive feedback

The positive feedback motif system consists of a single gene whose protein, when dimerized, enhances its own transcription and subsequent translation (see Figure 7(A)). For this system, the monomer concentration is x, the basal mRNA transcription rate is α, but when dimers of the protein bind to the promoter site the transcription rate increases to β. The positive feedback motif has been analyzed in various ways [18, 83, 4, 20], but the question we are considering here is new. What is the effect of changing the copy number (the number of motifs) on the steady state behavior of the system?

In the positive feedback motif considered here, dimerization precedes regulated recruitment. When $\mathcal{N}$ is below some threshold there will be insufficient concentration of dimers to enhance transcription. Thus, transcription will occur predominantly at basal levels. However, for copy numbers above some threshold the coupling between motifs will lead to enhanced transcription at activated levels. Hence, the steady state gene expression will jump nonlinearly as a function of $\mathcal{N}$. The effect of copy number on expression dynamics is depicted in Figure 8. Analytical predictions of a nonlinear jump in the steady state gene expression are confirmed by numerical simulation (see Figure 9(A), where the threshold is $\mathcal{N} = 3$ for the parameters considered). This jump in steady state expression can be formally explained due to the presence of a

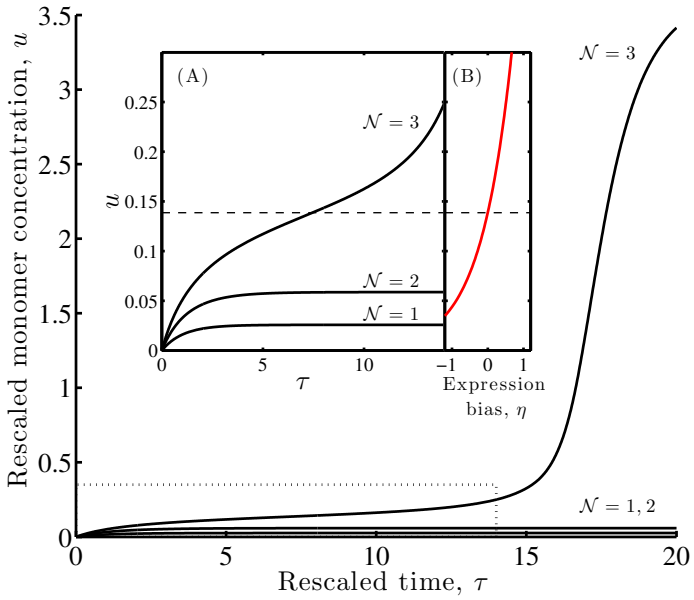

Figure 8: Dynamics of a positive feedback loop (see Equation (83) in Appendix A for detail) with re-scaled translation rates $\hat{\alpha} = 0.025$, $\hat{\beta} = 1.3$ given copy numbers $\mathcal{N} = 1, 2$, and 3. The inset has two parts: (A) Blow-up of the rescaled concentration dynamics, $u(\tau)$, for the rescaled time $0 \leq \tau \leq 14$; (B) Concentration u vs. expression bias $\eta = \log_{10}(\hat{\beta}u^2/\hat{\alpha})$. The expression bias represents the relative frequency of activated and basal transcription. The dashed line denotes the value of u such that the rate of basal transcription is equal to the rate of activated transcription, i.e. $\hat{\beta}u^2 = \hat{\alpha}$. Note that expression bias is on a log-scale and denotes the relative contribution of activated vs. basal transcription, e.g.,when $\eta = 1$, activated transcription occurs at 10-times the rate of basal transcription.

series of saddle-node bifurcations in which an alternative stable state (with high gene expression) appears in the system and eventually dominates the original (low gene expression) state.

We cannot expect that all positive feedback motifs will lead to such a dramatic sensitivity to copy number variation. The condition under which the system exhibits dramatic sensitivity to CNV is very robust, $\beta > 9\alpha$, and is derived in Appendix A. Thus, so long as enhanced transcription is sufficiently greater than the basal transcription rate, changing the copy number may lead to a jump in the steady state monomer concentration. The value of the copy number at which the dramatic change of expression occurs is a tunable feature, that will have important consequences for

how the entire system responds to CNV. For example, the theoretical prediction of the critical threshold, $\mathcal{N}_c$ may be too low ($\mathcal{N}_c < 1$) or too high ($\mathcal{N}_c \gg 1$) to be of any biological relevance. In the former case, the system should always be dominated by activated transcription, and in the later case, the copy number may never reach a level where activated transcription is feasible.

3.2.3 Bistable feedback

The bistable feedback motif consists of two genes whose protein monomer concentrations we denote by x_1 and x_2, respectively (see Figure 7(B)). These two genes have overlapping promoters such that each bound dimer halts the transcription of the other gene and regulates its own. The basal transcription rates are $\alpha_{1,2}$ and the regulated transcription rates are $\beta_{1,2}$. Bistable feedback is characteristic of networks that involve switching between alternative gene expression states [33, 4]. Here, we are particularly interested in the case where one gene has higher basal transcription $\alpha_2 > \alpha_1$ but the other gene has higher regulated transcription $\beta_1 > \beta_2$. These conditions are motivated by studies of the genetic switch that regulates the decision between lysis and lysogeny within bacteriophage λ [2, 146, 97, 155, 193]. Based on previous analysis of the case where $\beta_2 = 0$ [193], we expect that changes in $\mathcal{N}$ can lead to drastic changes in protein monomer concentrations. However, it is not immediately clear what kind of changes will occur. Will both concentrations jump up or only one of them? Will any of the concentrations drop? We answer these questions by determining the relation between $\mathcal{N}$ and the steady state monomer concentrations (see Appendix A). For the bistable motif, it is possible that for small values of $\mathcal{N}$ there is only one steady gene expression state, which is a stable node. As $\mathcal{N}$ increases, the system will enter a bistable region in which alternative stable states corresponding to dominance by either of the two genes is possible. Finally, as $\mathcal{N}$ increases even further, the system will have a different steady state, again a stable node. This implies that

one gene will dominate at low $\mathcal{N}$ and the other gene will dominate at high $\mathcal{N}$. A graphical representation of the above observations can be obtained by plotting $\mathcal{N}$ along the x-axis and the ratio of the two concentrations, $\bar{u}_2/\bar{u}_1$, along the y-axis. In this case, it is apparent that the ratio is large at the first stable node, but drops down significantly at the second one (see Figure 9(B)). In biological terms, this means that the bistable feature of the network depends on copy number. For sufficiently low or high values of $\mathcal{N}$, the coupled set of motifs will have deterministic outcomes. At low $\mathcal{N}$, gene 2, with the higher basal transcription rate will dominate, whereas at high $\mathcal{N}$, gene 1, with the higher regulated transcription rate will dominate.

Analysis of the steady state behavior demonstrates that the above bifurcations occur when $\beta_1 > 9\alpha_1$ along with a second algebraic condition described in Appendix A. The first condition implies that enhanced transcription must be at least 9 times as great as basal transcription in one of the genes. The second condition is more complicated and involves transcription rates as well as protein degradation rates - the condition is satisfied for a wide range of parameters (see Appendix A). Thus, we obtain robust conditions for a copy number controlled genetic switch. The switch in abundance of regulatory proteins can lead to radically different phenotypic effects inside a cell or organism. For example, the fate of bacterial cells infected by multiple phages exhibit an acute sensitivity to changes in the multiplicity of infection [101, 97, 193]. In addition, the values at which bifurcations occur, the sharpness between alternative gene expression states, and other features are tunable by this copy number dependent effect.

3.2.4 Toggle switch

The toggle switch motif consists of two genes with different promoters such that the product of one gene inhibits transcription of the other [57]. Here we consider the general case in which each gene product dimerizes prior to binding and then partially

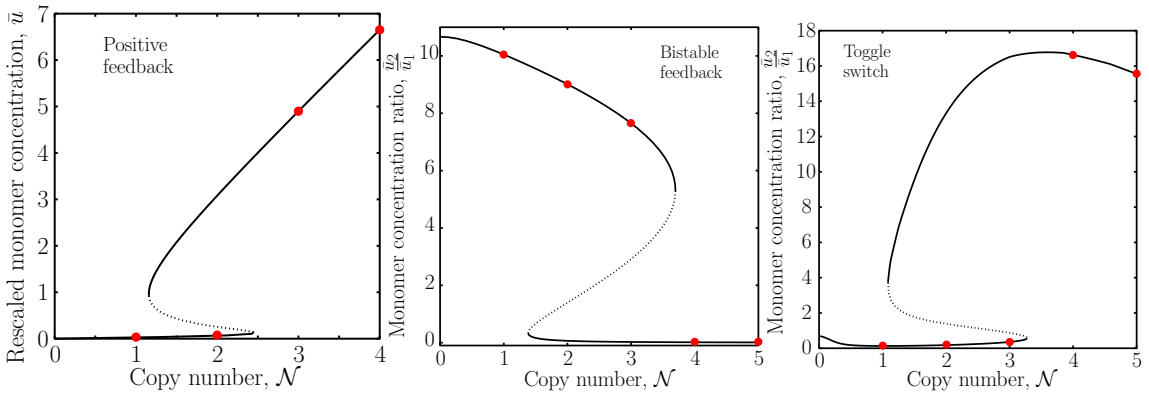

Figure 9: Steady state gene expression (or ratios) as a function of copy number, $\mathcal{N}$ for three motifs. In each case, solid lines are stable equilibria from theory, dashes lines are unstable equilibria from theory, and circles denote numerical simulation of the appropriate QSSA model. (A) Positive feedback: Numerically computed steady state monomer concentration as a function of copy number with $\hat{\alpha} = 0.025$, $\hat{\beta} = 1.7$. (B) Bistable feedback: relation between the copy number, $\mathcal{N}$ and the ratio of the steady state monomer concentrations; $\hat{\alpha}_1 = 0.063$, $\hat{\beta}_1 = 1.5$, $\hat{\alpha}_2 = 0.67$, and $\hat{\beta}_2 = 0.0083$. (C) Toggle switch: Relation between the copy number, $\mathcal{N}$ and the steady state monomer concentrations; $\hat{\alpha}_1 = 6.3$, $\hat{\beta}_1 = 0.14$, $\hat{\alpha}_2 = 4.3$, and $\hat{\beta}_2 = 0.43$.

inhibits transcription of the other, not completely. The basal transcription rates are $\alpha_{1,2}$ and the repressed transcription rates are $\beta_{1,2}$ where $\beta_{1,2} < \alpha_{1,2}$. The schematics of this system is shown in Figure 7(C). The motifs we have analyzed thus far share a common feature – positive feedback loops. Moreover, we saw that the conditions that guarantee existence of essential bifurcations in the two motifs would not be satisfied without positive feedback. So, how does the copy number affect behavior of genetic networks with only negative regulation?

For the toggle switch motif, there is a single steady state for low values of the copy number. As $\mathcal{N}$ increases two consecutive saddle-node bifurcations occur, first creating a new stable node and a saddle and then colliding the saddle with the old stable node. We also see that the steady state concentration ratio, $\bar{u}_2/\bar{u}_1$, is small before the first bifurcation and large after the second one (see Figure 9(C)). Biologically speaking this means that the dominant gene in a toggle switch can depend on the copy number

of the motif.

Analysis of the steady-state expression shows that existence of the two bifurcations and the resulting jump in the ratio of the steady state concentrations are quite robust with respect to the parameter values (see Appendix A). Thus, even without positive feedback a genetic network may switch to a drastically different state as the copy number changes. Note that the toggle switch of Gardner and Collins [57] corresponds to the case of complete mutual halting of transcription ($\beta_i = 0$). The Gardner and Collins switch exhibits behavior significantly different from the case $\beta_i > 0$. While at small values of $\mathcal{N}$ there is still a single stable node, increasing the copy number will lead to only one saddle-node bifurcation. At larger copy number, two gene expression states are possible, corresponding to dominance by either gene respectively. Hence, increasing copy number leads to bistable behavior, in which steady-state outcomes depend on initial conditions and the strength of stochastic effects.

3.2.5 Repressilator

The repressilator motif consists of three genes with a circular network structure such that gene 1 represses gene 2, gene 2 represses gene 3, and gene 3 represses gene 1 (see Figure 7(D)) [52]. Unlike the previous motifs, it is known that the repressilator can exhibit sustained oscillations. The corresponding stable limit cycle emerges from a stable node via a supercritical Hopf bifurcation [176, 52, 20]. What we are interested in is whether changes in copy number can switch the system between a single steady state and sustained oscillations.

In the repressilator motif, as before, we consider the situation where dimerization precedes binding to promoters. As we show in Appendix A, the above system has a single symmetric steady state where each of the three protein concentrations are at identical levels. Analysis of this steady state reveals that a Hopf bifurcation occurs when the copy number passes a critical threshold. A numerical simulation, shown in

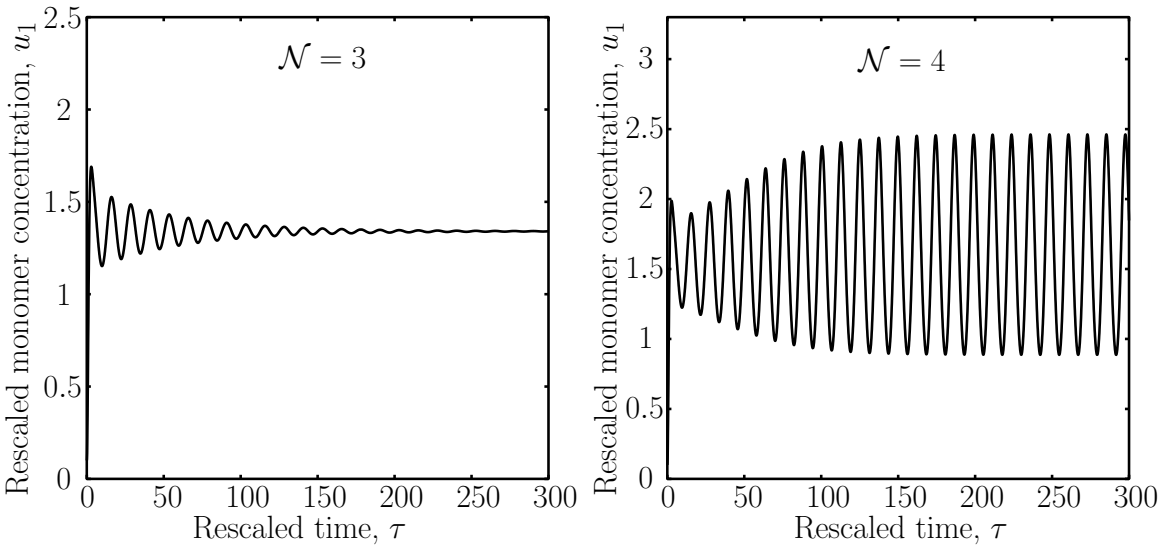

Figure 10: An example of the onset of oscillations in the repressilator as the copy number changes. Here $\hat{\alpha} = 1.25$, rescaled degradation of proteins is $\hat{\gamma} = 0.80$, and the critical value for the onset of oscillations as predicted in Appendix A is $\mathcal{N} = 3.43$.

Figure 10, confirms the above finding. Biologically speaking, when copy number is low, the circular series of transcriptional feedbacks is insufficient to allow dominance by a single gene in time. Increases in copy number allow a single gene to dominate for a short period, followed by the rise of its inhibitor and so on. Thus, at a sufficiently large copy number, the repressilator can exhibit oscillatory behavior. Copy number itself acts as a proxy for the degree of coupling in this system. We should keep in mind, however, that the estimate of the bifurcation point where the switch from steady to oscillatory behavior occurs is, in fact, a rather crude one deriving from the use of the QSSA. Nevertheless, it does demonstrate that changing the copy number can not only drive a genetic network to a different state, but also make it oscillate, and that the threshold of oscillations is a tunable quantity.

3.3 Discussion

We have demonstrated that copy number is a key control parameter in the expression dynamics of simple network motifs. Changing the copy number can make a network

switch to an entirely different equilibrium gene expression state as well as move it to and from an oscillating regime. Our results stand in contrast to previous assertions that target gene expression is proportional to gene copy number [21, 64]. Gene expression can be nonlinearly related to gene copy number because of feedbacks found in even the simplest of network motifs. Such nonlinearities are found even when the balance among gene components is maintained. Though not every small-scale CNV will lead to large-scale changes in gene expression, we have found a set of principles to understand when such a link occurs. In the cases of positive feedback, bistable feedback, and toggle switch motifs, we are able to find general conditions for the presence of qualitative sensitivity to copy number. In more technical terms, we have solved for the sufficient conditions for the existence of saddle-node bifurcations within a set of nonlinear dynamical systems. This has dramatic consequences for systematic analysis of the emergence and maintenance of CNV.

Importantly, our findings hold despite significant variation in parameters values associated with the molecular details of regulated recruitment (see Appendix A). Thus, sensitivity of motifs to CNV applies to a broad range of cellular contexts. The robustness of genetic regulatory networks to noise [99, 141] and gene duplication [188, 35] have been highlighted. Our findings suggest that there are limits to robustness, particularly with respect to gene duplication. The bifurcation conditions we derived for each motif provide guidance as to the range of kinetic parameters in which network fragility may be expected.

There are many challenges remaining in the study of the link between CNV and phenotypic effects. The networks we have considered are small components of complex gene regulatory networks. It remains an open question whether and to what extent these results scale up to larger, more complex networks [4, 44, 68]. For example, how have actual networks evolved with respect to the critical values of copy number which can lead to qualitative shifts in system behavior? Though we have studied

the effect of varying the copy number of motifs, it is worthwhile to examine the effects of copy number imbalances in complex motifs. Note that in this chapter we have assumed fully coupled network motifs, whereas the dynamics of intracellular transport of regulatory elements is certainly more complex [166, 155]. There are a number of areas in which we believe further examination is likely to yield successes in applying the theory presented here: host-phage dynamics, synthetic biology, and evolution via gene duplication. We discuss each of these areas below.

First, in the case of host-phage dynamics, there may be selection pressure favoring sensitivity to copy number, as in the case of temperate viruses whose exploitation strategy depends on the multiplicity of infection [101, 97]. In Chapter 2, we demonstrated that the number of phage DNA copies inside a bacterial cell has a dynamical effect on the decision making circuit of bacteriophage λ. Hence, coinfecting phages can in principle make collective decisions about a cell's fate. A small number of viruses can direct regulatory machinery toward lysis, whereas the coinfection of a single host by many viruses leads to a latent infection. Different phages differ in their response to co-infection, and so the response to coupling decision modules is likely to be an evolvable feature of phages' life histories. An alternative hypothesis for the link between cell fate and multiple infection is that each injected phage genome experiences a distinct micro-environment [155]. Even in such a case, coordination of phage response depends on synchronization of decision modules, though perhaps on different time scales.

Next, of relevance to synthetic biology, CNV may alter dynamics of gene regulatory networks that have been engineered *de novo* or modified to acquire new functions [54]. Here, we briefly discuss two experimental studies in which qualitative changes in gene expression were observed in synthetic networks as a consequence of small scale changes in the copy number of gene regulatory components. In one case, an *E. coli* gene regulatory circuit was designed to exhibit both sustained oscillations

and toggle switch behavior [9]. The copy number of a key activator module in the circuit (controlled by the *glnAp2* promoter) was increased by inserting it closer to the origin of replication. Comparison of gene expression showed a 20% decrease in the degree of damping of oscillations when the activator was located near the origin as opposed to near the terminus. In another case, a re-engineered budding yeast pheromone response pathway was designed to exhibit bistable response to pheromone induction [82]. Bistability depended sensitively on the number of positive feedback modules inserted into the yeast cells. A minimum of three tandem copies of the $P_{FUS1J1} - STE11\Delta N$ construct was necessary for a sustained positive feedback response, whereas one or two copies did not lead to a sustained response. Though these are only two examples, they both suggest that experimental studies of the sensitivity of small genetic circuits to CNV may be necessary if regulatory motifs are to be used as reliable building blocks of more complex networks [68].

Finally, gene duplication is considered to be a major factor in the evolution of novel phenotypes. According to the theory of neofunctionalization, duplicated genes are initially redundant, and, on occasion, one of a duplicate pair may diverge to perform some new function [111]. In fact, the number of retained gene duplicate pairs is unexpectedly high [111], with extensive experimental evidence that duplicate genes retain functional compensation over long periods of time [46, 47, 132]. Duplicated genes or motifs may not be strictly redundant, even initially. The evolution of network motifs subsequent to duplication may depend on global network context [145]. In the current theoretical framework, it is apparent that an extra copy of a gene or motif caused by a duplication event can lead to a shift in expression past some functional threshold. Thus, a new feature could emerge immediately, augmenting or modifying previous function. The possibility that duplicated genes are not redundant is supported by a number of evolutionary studies [76, 66]. This is not to say that large-scale gene expression given a gene duplication event must be the norm. To the

contrary, if the effect of an extra copy was somehow buffered, then the present dynamical framework of gene regulation would be consistent with a model of evolution via neofunctionalization.

These three biological examples reflect a small fraction of ongoing research by scientists from many disciplines to understand how CNV impacts a broad range of biological phenomena. Though our treatment of gene regulation is closest to the mechanisms of regulated recruitment within bacteria and viruses, we envision that a copy number effect may be present from viruses to higher eukaryotes. This effect may have as its hallmark, a dramatic change in gene expression given a small change in copy number. Even if such a dramatic change represents the exception in gene regulatory networks, when such a change does occur it may have exceptional implications in modifying biological function. Whether in the case of genomic structural variation in humans or bacteriophage infections, variation in copy number is ubiquitous. At minimum, we hope to have provided some first steps toward constructing quantitative models of regulated recruitment that take into account CNV.

3.4 Methods and Models

3.4.1 Modeling framework

There are four steps involved in our quantitative approach to linking copy number variation with gene expression dynamics: (Step 1) Define the molecular processes involved in the gene regulatory network motif; (Step 2) Convert motif into a mathematical language of gene regulatory dynamics; (Step 3) Simplify the mathematical model using a series of quasi-steady state approximations (QSSAs) [20]; (Step 4) Solve for the steady state gene expression as a function of copy number. A schematic of Steps 1-3 can be found in Figure 7. For each network motif we model the following molecular processes: transcription of mRNA, translation of mRNA into proteins,

dimerization of monomers into dimers, de-dimerization, binding of dimers to promoters upstream of genes, unbinding of dimers from promoters, degradation of mRNA, and degradation of proteins. Each of these processes is assumed to obey simple mass-action kinetics with corresponding kinetic rates such that any particular network motif can easily be transformed, in Step 2, into a nonlinear dynamical system [4]. In Step 3, concentrations within the QSSA model are described in terms of the slowly varying monomer concentration. Importantly, the QSSA model retains the equilibrium values of the original model and is analytically tractable. In Step 4, we are able to find the copy number dependence of gene expression via analysis of the QSSA model confirmed by computer simulation (see Appendix A).

CHAPTER IV

TO LYSE OR NOT TO LYSE: TRANSIENT-MEDIATED STOCHASTIC FATE DETERMINATION IN CELLS INFECTED BY BACTERIOPHAGES[1]

4.1 Introduction

Biochemical pathways and feedbacks in gene regulatory networks (GRNs) shape when and how much genes are expressed. Differential gene expression can lead to qualitative changes in cellular phenotypes, whether via alternative cell fate determination in unicellular organisms (e.g., competence [168], sporulation [174], persistence [15], and infected cell fate [146]) or via cell differentiation in multi-cellular organisms (e.g., lineage determination [129]). The steps leading to qualitative changes in phenotype are not strictly deterministic. Gene regulation is an inherently noisy process involving transcription control, translation, diffusion and chemical modifications of transcription factors, all of which may be characterized by stochastic fluctuations due to low copy numbers of regulatory molecules [118, 89, 90]. As a result, genetically identical cells can have marked differences in the state of regulatory molecules even when faced with identical environmental conditions [170, 53, 113]. Explanations for alternative cell fate determination generally presume the existence of multiple stationary states within the GRN [112, 110]. Determination of cell fate is therefore usually described as the result of the interplay between noise and deterministic dynamics of GRNs which determines the relative frequency of each decision [199, 112].

A potential problem with this explanation is that cellular decision making occurs

[1]This chapter appeared in the PLoS Computational Biology [87].

within finite time. From a theoretical point of view, differences in asymptotic dynamics are not necessary for regulatory dynamics to reach markedly different transient states. The hypothesis that transient dynamics can drive cell fate determination has been suggested in the context of HIV-1 latency where a bistable response is observed despite the purported monostability of the GRN [192]. Here, we take a generalized approach to a similar problem by considering cell fate determination as the result of stochastic transient dynamics of a GRN. Our starting point is the fact that extrinsic variation can drive substantial differences in the transient state of regulatory molecules [4]. That is to say, ensembles of cells with the same initial state of regulatory molecules which are exposed to two different conditions can follow distinct transient trajectories on average. In such a case, gene expression will be characterized by an early period in which transient trajectories are unresolvable with respect to the stochastic noise and a middle period in which they are markedly different. However, we claim that such transient differentiation in regulatory state need not be accompanied by marked differences in asymptotic, i.e., steady-state, behavior. Instead, we hypothesize that alternative cell fate decisions can be mediated by first passage processes of regulatory molecules [42, 148].

We examine the effect of first passage processes in stochastic GRNs within the initial decision switch between lysis and lysogeny by phage λ. Bacteriophage λ is perhaps the simplest example of an organism with alternative developmental modes, which are quiescent (lysogenic) and productive (lytic) growth upon infecting *E. coli* cells [73, 146, 49, 137, 40]. Here we focus on how λ-phage-infected cells are lysed or become lysogens as a function of the number of coinfecting phages (also known as the cellular multiplicity of infection denoted as $\mathcal{M}$). Experimental infection assays have revealed that *E. coli* cells that are multiply infected tend to become lysogens whereas singly infected cells tend to be lysed [101, 97]. The decision to lyse a cell or enter lysogeny is stochastic [8, 165], and the fraction of lysogeny is a probabilistic

function of the number of coinfecting phages and cell volume [171, 193, 201]. Cells that become lysogenic may later spontaneously induce leading to virion production and cell lysis. The stability of the lysogenic state has also been evaluated in light of first exit problems [11], many of whose concepts we adapt in the current model of the initial decision switch.

A significant advantage in modeling phage λ is that the core pathways of lysis-lysogeny have been studied extensively. Subsequent to infection, the repressors (CIs) bind cooperatively to adjacent operator sites [77], and cooperative binding can induce DNA loops which enhance the stability of the lysogenic state [63, 13, 128]. Early quantitative studies of the initial lysis-lysogeny decision utilized statistical thermodynamic models which described the dynamics of gene regulation by cooperative binding of CI [2, 163]. Arkin *et. al.* developed a fully stochastic model based on transcription, translation and protein interactions [8]. Whether cells were fated to lysis or lysogeny was ascribed to intrinsic stochasticity, whose complexity rendered it intractable for mathematical analysis. More recently, theoretical work has suggested that alternative decisions of lysis and lysogeny may be due to inherent bistability of the phage λ GRN with respect to changes in copy number concentration ($\mathcal{M}$ divided by host cell volume) [193]. However, this model presumes that differences in asymptotic dynamics lead to changes in cell fate, without considering stochastic effects in transient dynamics.

In this study, we demonstrate that biased alternative cell fate decisions are possible due to transient divergences within gene regulatory dynamics. As evidence, we develop and analyze a quantitative model of a GRN of phage λ based on empirical analyses of viral infection. Although the structure of the phage λ GRN is relatively well established, the quantitative values of most kinetic parameters involved in viral gene regulation remain either unknown or poorly constrained. We examine two sets of kinetic parameters close to consensus empirical estimates which we refer to

as transiently divergent and asymptotically divergent, respectively. We show that the dynamics of the GRN with these parameter sets are similar shortly after phage infection but the asymptotic dynamics are qualitatively distinct as a function of viral genome concentration. Next, we compare the fraction of lysogeny as a function of viral genome concentration in the two parameter sets. Cell fate is determined via first passage processes of two regulatory proteins, Q and CI, corresponding to lysis and lysogeny, respectively (see Section 4.4). We find that equivalent responses of cell fate to changes in viral genome concentration can be obtained with either parameter set, suggesting caution must be applied in interpreting alternative cell fate determination as a hallmark of bistability. In the process, we also discuss how thresholds of first passage processes can change the fraction of lysogeny and the time scale of decisions. Finally, we compare model results with experimental data on cell fate outcomes from single cell assay [201]. We propose an alternative data collapse of the observed cell fate outcomes, consistent with a previously unidentified gene dosage compensation mechanism. We show that including gene dosage compensation at the mRNA level in our stochastic model of transient fate determination also leads to the form of data collapse observed in the single cell study. We conclude by discussing means to reconcile multiple competing hypotheses for observed heterogeneity in the phage λ GRN.

4.2 Results

4.2.1 Deterministic dynamics of qualitatively identical phage λ decision switches can be asymptotically or transiently divergent

We first analyze a deterministic model of a GRN of phage λ (see Figure 11, and Equation (63) in Section 4.4 for detail). Prior to phage λ infection, there are no viral proteins and mRNAs in the host cell. A cell can be infected by $\mathcal{M}$ phages, which we vary one to five for a fixed cell volume. Cell fate, either lysis or lysogeny, is determined based on the first passages of a pair of fate-determining regulatory molecules, CI and

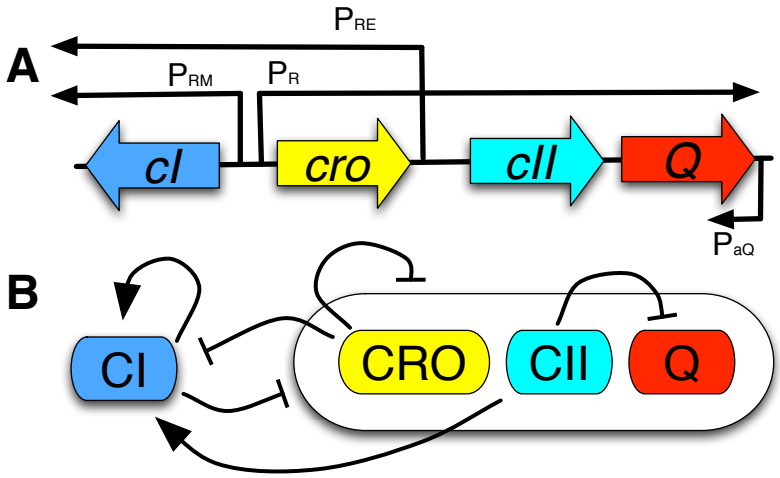

Figure 11: Core genetic components of lysis-lysogeny decision switch in phage λ. (A) Schematic diagram of genes and promoters. CI and CRO dimers are the transcription factors for P_{RM} and P_R while P_{RE} and P_{aQ} is controlled by CII tetramers. Black arrows represent open reading frames of promoters when activated (R_R, P_{RM} and P_{RE}) and antisense transcript aQ. (B) Interactions among gene products. Regular and blunt arrows represent positive and negative feedbacks, respectively. CI dimers are self-activators while repressing the other genes, and CRO dimers repress all the genes in the system. CII tetramers activate cI transcription, and suppress Q expression by transcribing antisense mRNAs.

Q (see Figure 12(B,E)). We model lysogeny as occurring when CI exceeds a concentration threshold and lysis as occurring when Q exceeds a concentration threshold. We set the value of these thresholds at 100 nM each, and explore the impact of varying these thresholds levels. Values of kinetic parameters necessary for modeling the lysis-lysogeny decision switch are known to within a few percent error in some cases, unknown in other cases, or have estimates with significant uncertainty (see Table 2). We chose two sets of parameters which are close to the consensus estimates, but that show markedly distinct asymptotic behaviors especially when $\mathcal{M} = 1$. GRNs with these two sets are asymptotically and transiently divergent, respectively (Figure 12). We define a phage λ GRN with a set of kinetic parameters to be asymptotically divergent if each deterministic trajectory for $\mathcal{M} = 1,\ 2, .., 5$ crosses the CI and Q thresholds only once. Otherwise, a GRN is referred to as transiently divergent.

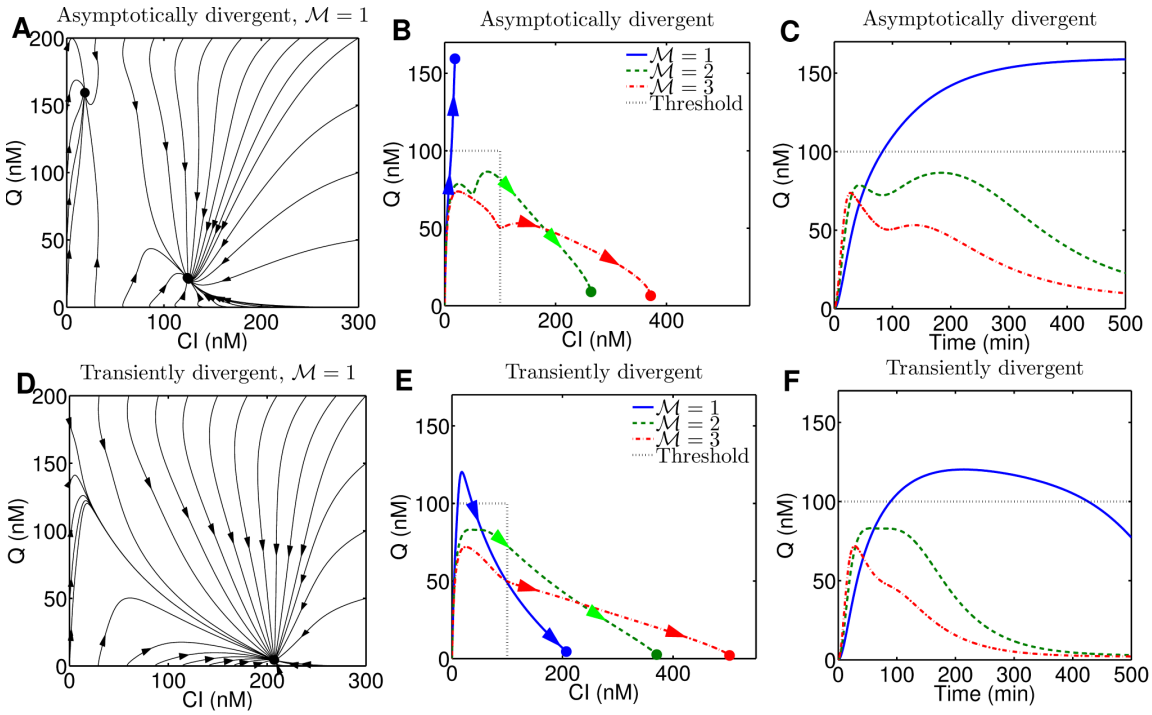

Figure 12: Dynamics of regulatory proteins, CI and Q, when the GRN is asymptotically divergent and transiently divergent. (A) Phase diagram of CI-Q dynamics when asymptotically divergent for $\mathcal{M} = 1$. Note that the system is bistable. (B) Phase diagram of CI-Q dynamics starting from no viral proteins when asymptotically divergent. Thresholds of CI and Q (both at 100nM) represent the concentrations above which decisions are lysogenic and lytic, respectively. Trajectories cross the threshold only once. (C) Asymptotically divergent dynamics of Q concentration as a function of time. (D) Phase diagram of transiently divergent system with $\mathcal{M} = 1$. Note that the system is not bistable. (E) Phase diagram of CI-Q dynamics of the transiently divergent phage λ GRN. At $\mathcal{M} = 1$ the deterministic trajectory crosses the threshold three times, and decisions change from lysis to lysogeny as a function of time. (F) Transiently divergent Q dynamics.

The transient dynamics for the phage λ GRN given either parameter set (either asymptotically or transiently divergent) are similar during the time scale of lysis-lysogeny decision ($<$ 100min). The asymptotically divergent phage λ GRN exhibits lysis for $\mathcal{M} = 1$ and lysogeny when $\mathcal{M} > 1$. Note that the ratio between CI and Q changes dramatically as a function of $\mathcal{M}$ from having far more Q to having far more CI at steady state. Only when $\mathcal{M} = 1$ are there two possible steady states, but the initial condition leads to lysis (Figure 12(A)). In contrast, the transiently divergent

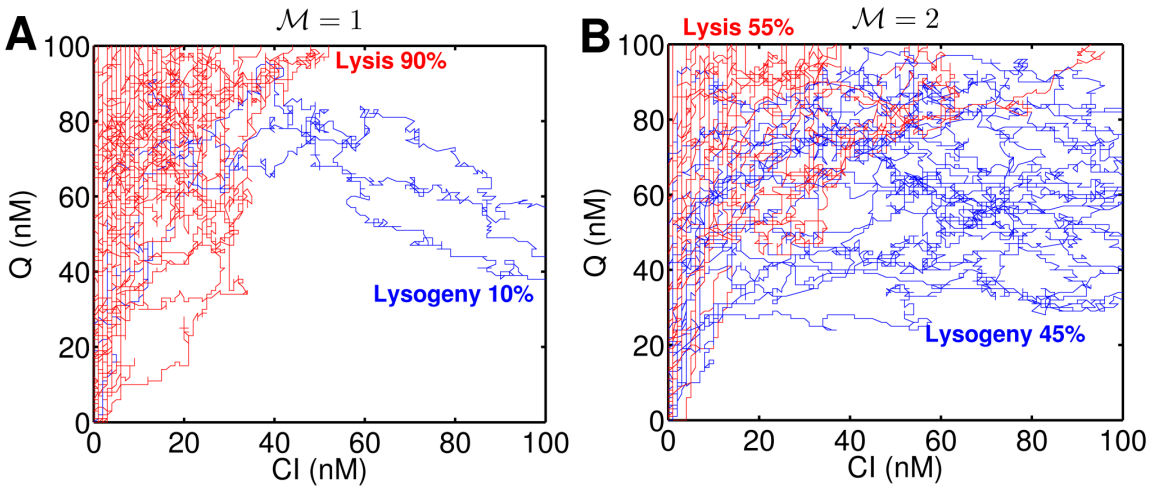

Figure 13: Stochastic realization of C and Q dynamics for (A) $\mathcal{M} = 1$ and (B) $\mathcal{M} = 2$. Trajectories are sampled for every 1/4 minute. The system is transiently divergent, and thresholds are set at 100nM for both CI and Q. Each curve represents a single realization, and 50 realizations are shown here. Red trajectories indicate that decisions are lytic whereas blue ones represent lysogeny.

GRN is monostable for all values of $\mathcal{M}$ that we considered. Further, the steady-state CI and Q concentrations have far greater levels of CI than Q, suggesting that an asymptotic analysis would suggest that the transiently divergent GRN would always lead to lysogeny. However, note that when $\mathcal{M} = 1$, Q increases rapidly, exceeds the threshold for lysis, and only later does it drop down and approaches a case where Q is low and CI is high (Figure 12(D-F)). Thus, there is an inconsistency between expectations for cell fate determination as viewed in finite time vs. that viewed asymptotically.

4.2.2 Alternative cell fates as determined by transient viral gene regulation

The initial lysis-lysogeny decision of phage λ is sensitive to the external conditions of $\mathcal{M}$ and cell size. Empirical analyses have shown this decision to be highly stochastic with the fraction of lysogeny between 20% and 90% for physiologically relevant $\mathcal{M}$ and cell size [201]. To model the stochastic nature of this decision, we assume that

first passage processes of CI and Q determine whether lysis or lysogeny occurs in an infected cell. Lysogeny occurs if CI reaches its critical concentration before Q does. Lysis occurs if the opposite holds true. We follow the approach of Arkin *et. al.* [8] and run fully stochastic simulations of the phage λ GRN while setting both lytic and lysogenic thresholds at 100 nM (see Section 4.4). We assume that reaching a decision of lysis or lysogeny brings a topological change to the GRN. Thus, we stop the dynamics at the time of a decision since our phage λ model cannot describe the post-decision regulatory dynamics. Figure 13 depicts a subsample of trajectories in the phase space of CI-Q labeled according to which decision is reached via a first passage process. Note that there is a delay for CI to be expressed since sufficiently abundant CII is required for initial CI expression. In contrast, Q can be produced immediately after phage infection. When the host is singly infected, lysis is the dominant decision, and CI does not build up until a significant amount of Q is produced (Figure 13(A)). At higher $\mathcal{M}$ ($\mathcal{M} = 2$ for Figure 13(B)), CII and Q are produced at a higher rate. Depending on the CII expression level Q can be repressed while CI becomes active which leads to lysogeny. In comparison to the deterministic dynamics described in the previous section, there is significant variability in the lysis-lysogeny bias of the GRN, though the bias itself is affected by changes in $\mathcal{M}$ and cell volume (as described in the next section).

4.2.3 Probability of lysogeny is an increasing function of phage genome concentration

We vary the volume of host cells (denoted as V) as well as $\mathcal{M}$ in order to investigate how cell fate responds to changes in the concentration of viral genomes ($\mathcal{M}/V$). For consistency with experimental studies and to model physiologically reasonable values, we vary $\mathcal{M}$ from one to five, and vary V from 0.5 to 2 μm^3. Figure 14 shows the fraction of lysogeny as a function of phage genome concentration. Regardless of bistability in the phage λ GRN, we find that first passage mediated decision making can lead to

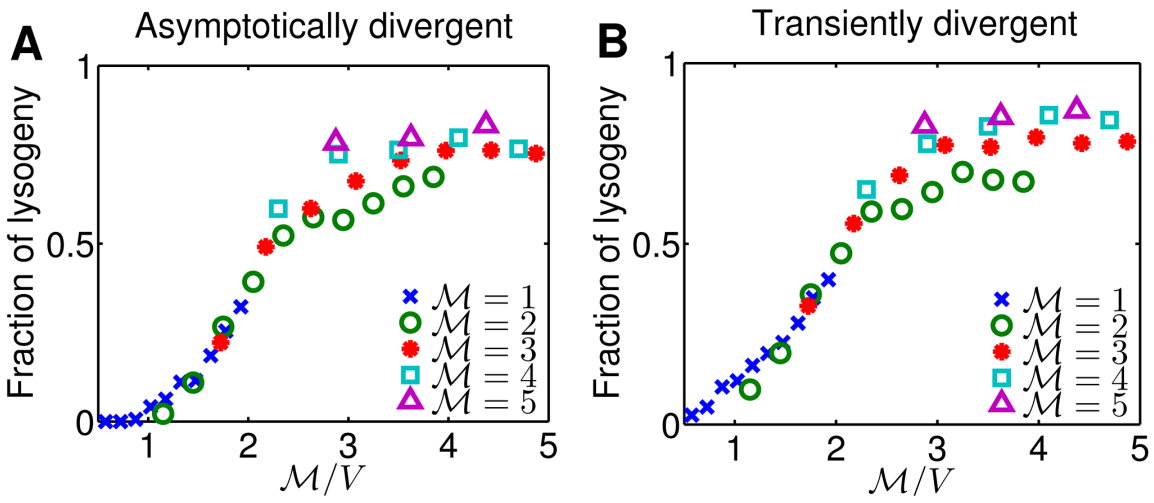

Figure 14: Response of phage λ to various phage genome concentrations when (A) asymptotically divergent and (B) transiently divergent. $\mathcal{M}$ and V represent the number of coinfecting phages and the host cell volume, respectively, so $\mathcal{M}/V$ is the phage genome concentration. Each point is the result from 5,000 simulations.

systematic biases in alternative cell fate determination. Phages preferentially enter lysogeny when multiple phages infect the same hosts while singly infected hosts tend to be fated for lysis. The relative frequencies of lysis or lysogeny can be collapsed as a function of an extrinsic parameter $\mathcal{M}/V$. Our results match the general trend of recent experimental observations which demonstrated that the fraction of lysogeny goes up as phage genome number increases or cell volume decreases [171, 201]. Importantly, the functional responses to phage genome concentration are nearly indistinguishable even for two parameter sets which have qualitatively different asymptotic dynamics (Figure 14 and Figure 12(B,E)). The biased decision response as a function of phage genome concentration is due to the similarity of transient dynamics, irrespective of asymptotic dynamics that could have been followed. Hence, the finding that infected cell fate can change from predominantly lytic (at $\mathcal{M} = 1$) to predominantly lysogenic (at $\mathcal{M} > 2$) is not necessarily a hallmark of an underlying bistable viral GRN nor of a bifurcation in the underlying dynamics as a function of $\mathcal{M}$ or $\mathcal{M}/V$. Despite

the agreement with prior empirical studies, note that our model does not predict systematic decreases in the lysogen fraction given a fixed value of $\mathcal{M}/V$ and increasing values of $\mathcal{M}$, as observed in a recent single-cell experimental study [201]. In the next section, we revisit the experimental data from Zeng *et. al.* [201] and in so doing, provide an alternative data collapse and a corresponding mechanism that is consistent with a modified version of the current stochastic model.

4.2.4 Mechanism of partial gene dosage compensation accounts for observed heterogeneity in lysis-lysogeny decisions

Zeng *et. al.* [201] measured the fate of multiply infected cells in which the number of phages and cell volume could be measured on a per-cell basis. The experimental protocol induces viral injection with an abrupt change in temperature and hence, infections are treated as simultaneous. The experimental data demonstrate that the fraction of lysogeny increases with viral concentration, $\mathcal{M}/V$ (Figure 15(A)). This trend agrees with prior experimental works showing that increases in co-infection number increases the likelihood of lysogeny [101, 97] and that increases in cell volume increases the likelihood of lysogeny [171]. However, there is significant amount of heterogeneity in the observed cell fate data other than strict dependence on $\mathcal{M}/V$ as suggested by theory [193].

In particular, Zeng *et. al.* [201] observed that the fraction of lysogens decreases with increasing $\mathcal{M}$ for a given ratio of $\mathcal{M}/V$. Zeng *et. al.* [201] suggested that the remaining heterogeneity in cell fate not explained by a strict dependence on $\mathcal{M}/V$ is due to a voting mechanism that takes place at the single-cell level. In this view, a unanimous decision of phages is required by phages for lysogeny [201] (presumably because a single phage that is fated to lysis would over-ride a decision by other phages for lysogeny). If each coinfecting phage is totally independent from each other, then one would expect the probability of lysogeny to be:

$$P_{lysg}(\mathcal{M}, V) = f\,(1/V)^{\mathcal{M}}\ , \tag{61}$$

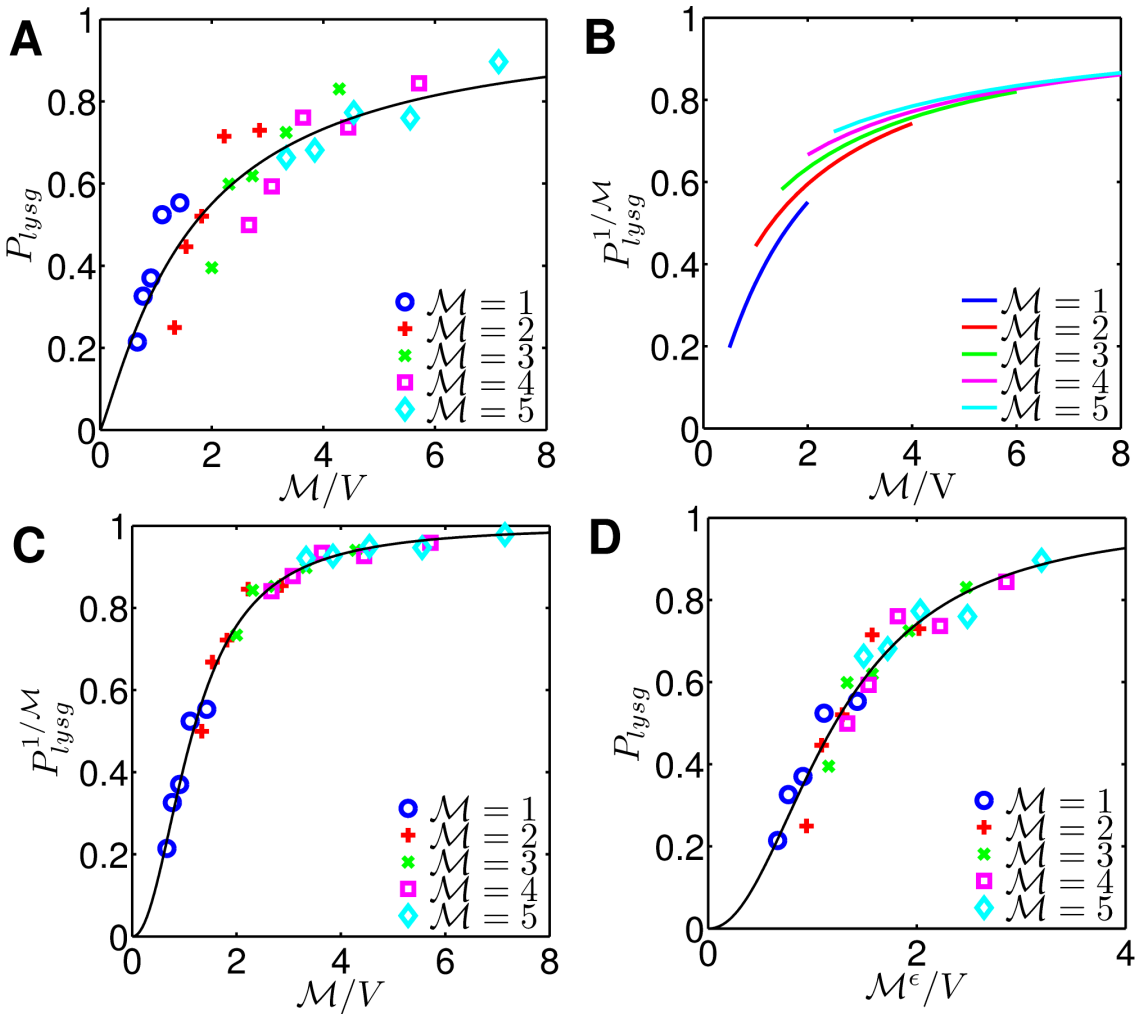

Figure 15: Alternative mechanisms underlying heterogeneity of lysis-lysogeny decisions. (A) Fraction of lysogeny plotted from single cell assays[201]. (B) Rescaled probability of $f(1/V)^{1,\mathcal{M}}$. Each phage within a host is completely independent from other phages, and decision of lysogeny becomes a function of host volume. Note that rescaled curves do not collapse into a single curve. (C) Rescaled probability of $f(\mathcal{M}/V)^{1/\mathcal{M}}$ proposed by Zeng *et. al.*[201] representing the probability of lysogeny for each individual infecting phage. Each phage independently "chooses" lysis or lysogeny. However, since the fraction of lysogeny for a single phage is a function of $\mathcal{M}/V$, phages sense the presence of other phages. Note that data from different $\mathcal{M}$-s collapse into a single curve. (D) Probability of lysogeny plotted against rescaled $\mathcal{M}^{\epsilon}/V$ when $\epsilon = 0.5$, corresponding to a mechanism in which gene expression from multiple copies is partially compensated. Due to partial dosage compensation, the transcription rate is not linearly proportional to $\mathcal{M}$, and the effective copy number is given as $\mathcal{M}^{\epsilon}$ where $0 \leq \epsilon \leq 1$. Note that the data from different $\mathcal{M}$-s collapse into a single curve. Black lines represent nonlinear curve fits into Hill functions.

where $f(1/V)$ is the probability that a cell of volume V infected by a single phage would become a lysogen. Figure 15(B) shows the fraction of lysogeny scaled with $1/\mathcal{M}$ power based on the empirical observations for the singly infected case. The re-scaled data for the five values of $\mathcal{M}$ should agree with $f(1/V)$ in an independent phage voting model. However, this rescaling does not form a single line. This suggests that there might be some inter-dependence between phages.

Indeed, the voting model proposed by Zeng *et. al.* [201] is actually a "quasi-independent" voting model. In this view, a unanimous decision of phages is required by phages for lysogeny [201]. However, the probability that any given phage decides for lysogeny becomes a function of the viral genome concentration, $\mathcal{M}/V$. Thus the fraction of lysogeny becomes

$$P_{lysg}(\mathcal{M}, V) = f_1 \left(\mathcal{M}/V\right)^{\mathcal{M}} . \tag{62}$$

where $f_1(\mathcal{M}/V)$ is the probability that a single phage reaches a lysogenic decision state given that it is in a cell of volume V with a total of $\mathcal{M}$ phages. The re-scaled probability of entering lysogeny at the whole cell level, $P_{lysg}^{1/\mathcal{M}}$, is shown in Figure 15(C). Notably, the re-scaled experimental data collapses on a single line, presumably $f_1(\mathcal{M}/V)$. Thus, this mechanism captures the characteristics of experimental data phenomenologically. However, the mechanism involves both independence and inter-dependence among phage genomes that remains un-identified at the subcellular level.

Here, we revisit the cell fate data of Zeng *et. al.* [201] and propose a mechanism of partial gene dosage compensation as an alternative explanation for the scaling collapse they observe. In this context, partial gene dose compensation means that a cell with multiple copies of a viral genome has smaller per-copy viral gene expression than a cell with a single viral genome. Indirect support already exists for this hypothesis. For example, Zeng *et. al.* [201] showed that the fraction of cells with halted growth increases with the number of co-infections, suggesting that viral genomes have adverse

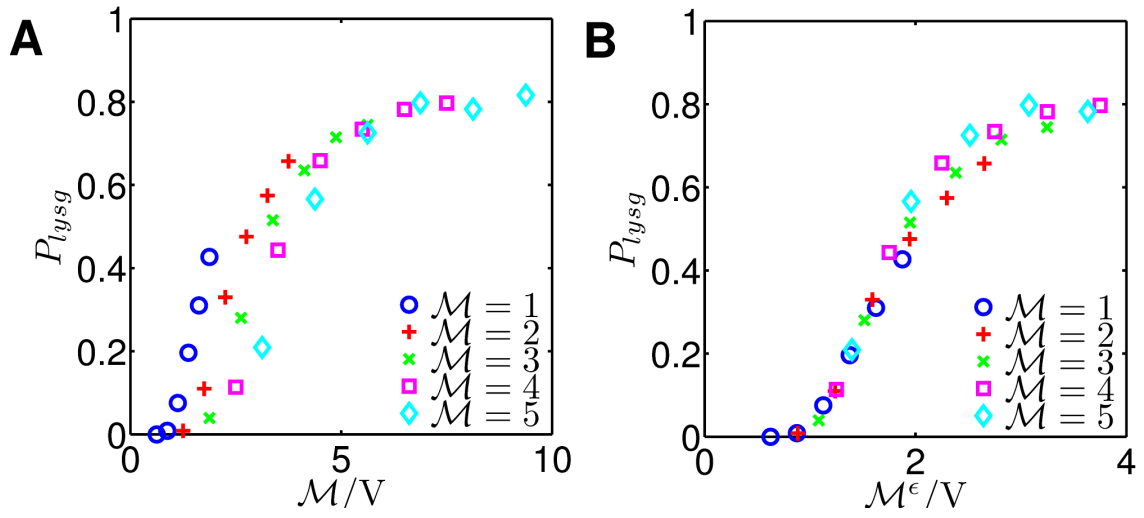

Figure 16: Effect of gene dosage compensation from stochastic simulations. (A) Fraction of lysogeny from stochastic simulations. Simulations with partial dosage compensation exhibit the nested pattern of $\mathcal{M}/V$ dependence as seen in the experimental data (see Figure 15). (B) Simulation results on the fraction of lysogeny from Figure 16(A) plotted with rescaled $\mathcal{M}^{\epsilon}/V$ when $\epsilon = 0.5$. The outcome of stochastic simulations with partial dosage compensation is consistent with experimental data (see Figure 15(A,D)). In this case, the GRN is asymptotically driven with CI and Q threshold at 100nM and 120nM, respectively, all other parameters are set according to Table 2 transiently divergent. Each point is the result from 3,000 simulations.

effects on cellular metabolism in addition to or instead of lysis. Earlier studies showed that phage λ infections repress host synthesis activity at the level of transcription [181] and translation [79]. The degree of repression depends on the number of coinfections, and more coinfections lead to greater repression. Broadly speaking, the mechanism (or mechanisms) underlying gene dosage compensation remains an open question. However, it has been widely noted that copy numbers of genes and chromosomes can differ among cells and individuals, but the resulting gene expression need not be a linear function of gene copy number [21, 178, 169].

Here, we assume that partial gene dosage compensation occurs at the level of transcription. Specifically, we assume that the total transcription rate of a gene is proportional to $\mathcal{M}^{\epsilon}$ where $0 \leq \epsilon \leq 1$ (see Section 4.4 and Equation (63)). ϵ is the quantitative measure of partial gene dosage compensation and RNA synthesis

repression by phage genomes. When $\epsilon = 0$, increases in viral genome have no effect on transcriptional rates, whereas when $\epsilon = 1$, transcriptional rates increase linearly with $\mathcal{M}$ (as in the original model described previously in this chapter). Hence, if a partial gene dosage mechanism is at work, then the lysogeny data should collapse when plotted against $\mathcal{M}^\epsilon/V$. Figure 15(D) shows the fraction of lysogeny against $\mathcal{M}^\epsilon/V$ which incorporates the effect of partial gene dosage compensation. Note that the data collapses into a single line, similar to the quasi-independent decision mechanism. The estimate of ϵ from experimental data is about 0.5, suggesting that the overall viral transcriptional activity in the host cells on a per-viral genome basis scales with $1/\sqrt{\mathcal{M}}$.

Hence, two distinct mechanisms: (i) quasi-independent decision making; and (ii) partial gene dosage compensation, can explain heterogeneous decision making from single cell assay based experiments using data collapse. Note that we cannot evaluate the quasi-independent mechanism using our model because doing so would require incorporating genome-specific changes (such as anti-termination events) or compartmentalizing the cell with respect to transcription and translation events (requiring even more unknown parameters than the current model). However, it is possible to explicitly incorporate partial gene dosage compensation in stochastic simulations (see Section 4.4). In brief, we modified transcriptional rates so that transcription increased with $\mathcal{M}^\epsilon$ instead of $\mathcal{M}$ and ran stochastic simulations with all other parameters as before. Figure 16 shows the fraction of lysogeny resulting from the stochastic fate determination model incorporating partial gene dosage compensation against $\mathcal{M}/V$ and rescaled $\mathcal{M}^\epsilon/V$. Stochastic simulations with partial dosage compensation exhibit the heterogeneous, yet strong dependence of lysogeny on $\mathcal{M}/V$. Moreover, the cell fate results of stochastic simulations collapse into a single line when $\mathcal{M}/V$ is rescaled as $\mathcal{M}^\epsilon/V$. Given the new scaling collapse, cells with the same $\mathcal{M}^\epsilon/V$ have a lower

chance of lysogeny given increasing values of $\mathcal{M}$, consistent with the pattern observed in the experimental study (Figure 15(D)). Hence, we propose that partial gene dosage compensation should be considered as an alternative mechanism to explain the heterogeneous cell fate of bacteria infected by bacteriophage λ.

4.3 Discussion

In this chapter, we have proposed and analyzed a transient mechanism of cell fate determination in terms of first passage processes of regulatory proteins. We applied this mechanism to the study of the initial lysis-lysogeny decision in bacterial cells infected by phage λ. We found that stochastic simulation of parametrized viral GRNs lead to changes in the the frequency of alternative fates for infected cells, either lysis or lysogeny, as a function of the genome concentration of infecting viruses. The biased response in cell fate outcome occurs despite intrinsic noise in the system and does not require the bistability of the underlying GRN. Hence, alternative and seemingly adaptive cell fate decisions may be due to transient divergence in stochastic trajectories of regulatory molecules and not necessarily due to underlying bistability. Finally, we showed that a partial gene dosage compensation is a candidate mechanism underlying noise in lysis-lysogeny decisions, as supported by both our quantitative model and experimental data.

Our central result is in contrast to the conventional perspective that multistability is required for alternative decisions [199, 110]. Multistability often requires cooperative binding as a necessary condition for the emergence of the two or more stable steady states in the GRN [57, 33]. A recent study showed that a switch system can arise in the absence of cooperative bindings [108]. Our study suggests that cooperative binding may occur and affect transient dynamics but not necessarily lead to bistability in asymptotic dynamics. Together these results suggest that GRNs which do not have bistability or cooperative bindings might be able to lead to alternative cell fate

determinations. Thus, it might be possible for a GRN to evolve (by natural selection) or to be designed (via synthetic means) to perform a complex task of alternative decision making in response to external stimuli without multistability. Note that such a transiently excitable GRN which differentiates transient and asymptotic phenotypes was experimentally demonstrated in *Bacillus subtilis* [177]. Generally, there exist examples of GRNs which are responsive to environmental signals and robust to changes of kinetic parameters [203] while other are sensitive to kinetic parameters. Sensitivity of transient dynamics to a GRN's kinetic parameters and thresholds might be a target of selection over evolutionary time scales. In this context, we examined how modifying thresholds for decisions can lead to systematic changes in lysis vs. lysogeny as well as decision times (see Figure 23 in Appendix B). The general result from the present analysis is that alternative determination requires separation of thresholds, which comes at the expense of slower decisions. Hence, transiently driven cellular decisions have the potential to be highly evolvable.

As we have detailed, stochastic simulations of the phage GRN proposed here can reproduce a number of characteristics for how the fraction of lysogeny changes with $\mathcal{M}$ and cell volume. Importantly, we find that lysogeny increases with increasing $\mathcal{M}$ [101, 97] and decreasing cell volume [171], and remains between approximately 20%-90% for physiologically reasonable values [201]. The bias in cell fate outcome in favor of lysogeny with increasing $\mathcal{M}$ may be adaptively significant. On average, high $\mathcal{M}$ implies that phages infect hosts frequently on the time-scale of decision-making and further, that phages are more abundant than their bacterial hosts. Lysis will further increase the phage-host ratio, and a previous study has speculated that phages seem to avoid depletion of hosts by entering lysogeny predominantly at high $\mathcal{M}$ [173]. However, if lysogeny is adaptively favorable at high $\mathcal{M}$, why is it that a small fraction of phages still enter the lytic pathway? The answer could be due to constraints in the resolvability of the GRN due to the strength of intrinsic stochasticity

in the GRN [201]. Or the stochasticity itself may be adaptive. Phages may have evolved to respond to changes in intracellular phage genome concentration in order to minimize the chance of extinction [12] by maintaining phage and lysogen population as a bet-hedging strategy [185]. Any such speculations require careful consideration of selective pressures imparted by ecological dynamics, game theoretic issues arising from co-infections by non-identical strains, and biophysical constraints and trade-offs arising at the intracellular scale [67].

However, the first set of stochastic simulations of the phage GRN presented in this chapter fail to predict the systematic decrease in the fraction of lysogeny given a fixed value of $\mathcal{M}/V$ and increasing values of $\mathcal{M}$ [201] (see Figure 14). We revisited the original single-cell data and demonstrated the existence of an alternative scaling collapse owing to a proposed partial gene dosage compensation mechanism. When we incorporate partial gene dosage compensation within our stochastic model, we are able to recover the alternative scaling collapse consistent with the empirical measurements of Zeng *et. al.* [201] (see Figure 15(D) and Figure 16(B)). What might cause partial dosage compensation to occur in multiple infected cells? In stochastic simulations here, dosage compensation is modeled explicitly at the transcriptional level, whereas in reality multiple factors can contribute to it, and may occur at both transcriptional and post-transcriptional levels. The degree of compensation might change depending on copy numbers of genes and chromosomes as well as other intracellular factors. Copy number variation (CNV) is common in biological organisms [142, 47], and previous studies suggested that gene expression can depend sensitively on CNV when uncompensated [124]. Indeed, one hypothesis is that gene regulatory networks have been selected for their lack of dosage sensitivity to avoid problems in gene expression that may arise when CNV occurs naturally [161]. Previous studies showed that phage λ represses overall activity of RNA and protein synthesis within infected

hosts depending on the number of coinfections [79, 181]. Viruses are known to control host cell cycle in eukaryotic cells [19], but how viruses affect the overall host transcriptional and translational activity in bacterial hosts is an open question. We believe that elucidating intracellular mechanisms of gene dosage compensation would be an important step toward understanding CNV and its resulting change in gene expression, at both the transient and steady state. In doing so, we also hope to provide a cautionary note: deducing explicit mechanisms from data collapses can be difficult, particularly when multiple data collapse schemes are consistent with observations.

In summary, this study proposed a novel intracellular decision-making mechanism to explain the variability in cell fate determination in multiply infected hosts. However, there can be other sources of variability underlying the lysis-lysogeny decision switch. First, the viral concentration, $\mathcal{M}/V$, in naturally infected hosts may be dynamic. Multiple phages infect a host sequentially, and a host can keep growing while being infected. Subsequent infections increase $\mathcal{M}$ over time, and infected cells may spend a substantial fraction of the time prior to cell fate determination with a value of $\mathcal{M}$ which is smaller than the final $\mathcal{M}$. Next, host cell growth decreases $\mathcal{M}/V$ whereas viral genome replication increases $\mathcal{M}/V$ during the infection cycle. Clearly the dynamic nature of viral genome concentration needs to be addressed even if experimental protocols have been designed to synchronize infections. Second, we assume the bacterial cytoplasm is well-mixed. Previous studies demonstrated that bacterial DNA, RNA and proteins have spatial patterns [164, 156, 182]. Bacteriophages are known to target cellular poles of hosts preferentially [51] which suggests phage genomes might be localized within bacterial cytoplasm. Hence, cell fate decision may be determined by local concentrations of regulatory proteins and quasi-independent cell fate determination by each virus. Finally, we assumed decision making as strict first passage processes arising from the consideration of thresholds

as absorbing states within a GRN dynamics. It is possible that decision making involves soft thresholds over which cells make decisions with some probability. There are studies which show duration of signals is critical to cellular decisions [117, 50], and there might be some minimum time interval during which the system is above a threshold to make a decision [115]. Even if experimental protocols can minimize the impact of one of these mechanisms, the evolution of the phage λ GRN would surely be impacted by all of them. Progress in identifying the importance of each of these issues at the molecular and evolutionary scales is relevant not only to the study of transient fate determination in phage λ, but to the study of cellular decision making in general.

4.4 Methods and Models

4.4.1 Gene regulation in phage λ

The fate of *E. coli* cells infected by phage λ are decided soon after infection by a set of so-called early viral genes [146]. Among them we consider four genes, *cI, cro, cII* and Q, and one antisense mRNA (aQ) (see Figure 11(A)). The expression of these genes are controlled by four promoters, P_R, P_{RM}, P_{RE} and P_{aQ}. P_R and P_{RM} share three operator sites which are targeted by CI and CRO. The natural form of CI is a dimer, and CI dimers act as self activators and repressors for other genes by binding to P_R/P_{RM}. CII tetramers can bind to P_{aQ} to transcribe aQ mRNA and P_{RE} to produce CI [139]. Dimers of CRO bind to P_R/P_{RM} to inhibit all the genes in the system (Figure 11(B)).

Immediately after phage infections there are no viral gene products. At this initial stage P_R is active which leads to an increase of CRO, CII and Q levels. If Q becomes sufficiently abundant, it will turn on genes which make progeny phages, and the infected host will be lysed. However, as CII concentration increases CII tetramers can activate CI transcription from P_{RE}, and CI expression level become further enhanced

by the positive feedback loop of CI at P_{RM}. CII also represses Q by transcribing aQ which facilitates Q mRNA degradation, and sufficiently high CI level leads to lysogeny [146]. Hence, lysis or lysogeny is determined based on which of either CI and Q reaches the threshold concentration first. When CI reaches its threshold, CI dimers begin to form tetramers and octamers which lead to DNA looping [150]. DNA looping is very stable while maintaining lysogeny and repressing genes which trigger lysis [128]. When Q reaches its threshold, a group of late genes responsible for making progeny phages will be turned on, and the host will eventually be lysed. Since translation occurs with a single protein at a time, simultaneous crossings of lytic and lysogenic thresholds are forbidden, and the decisions are mutually exclusive. In reality, decisions would not be triggered by infinitesimally short bursts over decision thresholds, but for simplicity we assume a decision is made when either CI or Q concentration reaches its threshold for the first time. The use of step functions instead of Hill function type responses has been used extensively in the study of quantitative gene regulatory networks [4]. Note that when phages multiply infect cells in natural settings, they do not do so simultaneously, and so $\mathcal{M}$ increases sequentially in time. However, for simplicity we only consider simultaneous coinfections, for which $\mathcal{M}$ becomes a parameter in determining cell fate rather than a dynamic variable. This choice of modeling simultaneous infections is also motivated by the the experimental protocol of Zeng *et. al.* [201] in which rapid temperature changes were used to synchronize phage infection of DNA into host genomes.

4.4.2 Quantitative model of phage λ decision switch

Here we express the interactions among *cI, cro, cII* and Q as well as aQ mRNA described in the previous section as a set of ordinary differential equations. If we apply quasi-steady-state approximation for dimers and tetramers, the system can be

described as

$$
\begin{aligned}
[cI\ mRNA] \quad \tfrac{dm_x}{dt} &= \frac{\mathcal{M}}{V}\alpha_x f_{RM}^{basal} + \frac{\mathcal{M}}{V}\beta_x f_{RM}^{act} + \frac{\mathcal{M}}{V}\delta_x f_{RE} - \gamma_m m_x\,, \\
[cro\ mRNA] \quad \tfrac{dm_y}{dt} &= \frac{\mathcal{M}}{V}\alpha_y f_R - \gamma_m m_y\,, \\
[cII\ mRNA] \quad \tfrac{dm_z}{dt} &= \frac{\mathcal{M}}{V}\alpha_z f_R - \gamma_m m_z\,, \\
[Q\ mRNA] \quad \tfrac{dm_Q}{dt} &= \frac{\mathcal{M}}{V}\alpha_Q f_R - \gamma_m m_Q - \zeta m_Q m_{aQ}\,, \\
[aQ\ mRNA] \quad \tfrac{dm_{aQ}}{dt} &= \frac{\mathcal{M}}{V}\delta_{aQ} f_{aQ} - \gamma_m m_{aQ} - \zeta m_Q m_{aQ}\,, \\
[\mathrm{CI}] \quad \tfrac{dX}{dt} &= \sigma m_x - \gamma_x X\,, \\
[\mathrm{CRO}] \quad \tfrac{dY}{dt} &= \sigma m_y - \gamma_y Y\,, \\
[\mathrm{CII}] \quad \tfrac{dZ}{dt} &= \sigma m_z - \gamma_z Z\,, \\
[\mathrm{Q}] \quad \tfrac{dQ}{dt} &= \sigma m_Q - \gamma_Q Q\,,
\end{aligned} \tag{63}
$$

where X, Y, Z and Q represent the total concentration of CI, CRO, CII and Q, respectively. $\mathcal{M}$ represents the number of coinfecting phages while V is the cell volume. m represents the mRNA concentration, and γ denotes the degradation rate where each subscript represent the species of associated gene/protein. Q and aQ mRNA become degraded by binding to each other and the adsorption rate is denoted as ζ. α, β and δ represent the basal, CI-mediated and CII-mediated transcription rates with subscripts indicating the species of mRNA. Note that α, β and δ are inversely proportional to V since the concentration change by a transcription event is proportional to $1/V$. We assume that the concentrations of dimers and tetramers are at quasi-steady states such as

$$
\begin{aligned}
X &= x_1 + 2x_2 = x_1 + 2c_d^x x_1^2\,, \\
Y &= y_1 + 2y_2 = y_1 + 2c_d^y y_1^2\,, \\
Z &= z_1 + 2z_2 + 4z_4 = z_1 + 2c_d^z z_1^2 + 4c_t^z c_d^{z2} z_1^4\,,
\end{aligned} \tag{64}
$$

Table 2: Parameters for transiently divergent and asymptotically divergent GRNs.

Parameter	Reference value	Reference	Asymptotically divergent	Transiently divergent
γ_x	$0.01(min^{-1})$	≈ 0 [66], 0.042 [25]	0.014	0.013
γ_y	$0.06(min^{-1})$	0.016 [67]	0.033	0.056
γ_z	$0.10(min^{-1})$	0.16 w/o CIII [68]	0.13	0.22
γ_q	$0.01(min^{-1})$		0.0095	0.016
γ_m	$0.1(min^{-1})$	0.12 [69]	0.1	0.1
α_x	$0.06(min^{-1})$	0.06 [36]	0.055	0.055
α_y	$0.84(min^{-1})$	0.84 [36], 3 [66]	0.82	0.70
α_z	$0.8(min^{-1})$	$<\alpha_y$	0.50	0.83
α_q	$0.75(min^{-1})$	$<\alpha_z$	0.78	0.63
β_x	$0.66(min^{-1})$	0.66 [36], 3.42 [70]	0.79	0.88
δ_x	$0.9(min^{-1})$	0.9 [25]	1.24	0.93
δ_{aQ}	$2(min^{-1})$		3.1	3.8
c_d^x	$0.05(nM^{-1})$	0.05 [71], 0.18 [72]	0.060	0.079
c_d^y	$5.8(nM^{-1})$	5.8 [73], 307 [74]	4.6	6.7
c_d^z	$0.05(nM^{-1})$		0.065	0.020
c_t^z	$0.05(nM^{-1})$		0.093	0.068
c_p^{aQ}	$0.2(nM^{-1})$		0.14	0.10
σ	$0.5(min^{-1})$		0.38	0.47
ζ	$0.1(nM^{-1}min^{-1})$	0.02 [75]	0.15	0.068
V	$1(\mu m^3)$	0.5 ~ 2.0		

where the subscripts 1, 2 and 4 represent the concentration of monomers, dimers, and tetramers of each respective protein. c_d and c_t are the dimerization and tetramerization constants, respectively. f_R, f_{RM}, f_{RE} and f_{aQ} in Equation (63) denote the probability of transcribable configurations for each promoters based on free energy change of possible states, and we follow the calculation of Shea and Ackers for $f_{RM}(x_2, y_2)$ and $f_R(x_2, y_2)$ [163] and Arkin *et. al.* for $f_{RE}(z_4)$ [8] (see Appendix B). f_{RM} has two modes of transcription denoted as basal and activated depending on (x_2, y_2). Response of f_{aQ} is a first order Hill function which is

$$f_{aQ}(z_4) = \frac{c_p^{aQ} z_4}{1 + c_p^{aQ} z_4}\,. \tag{65}$$

For stochastic simulations, we chose two parameter sets which lead to a transiently and asymptotically divergent lysis-lysogeny decision switch. Parameter values for the transiently divergent and asymptotically divergent cases are listed in Table 2. To calculate the fraction of lysogeny, we used 5,000 realizations of the stochastic model.

Our simulations are based on Equation (63) and are fully stochastic as implemented using the Gillespie algorithm [61] (see Appendix B for details).

4.4.3 Modeling gene dosage compensation

When gene dosage is compensated, the effective copy number, which is the fold change of transcription rate, is smaller than the actual copy number. Here we assume the effective copy number scales as $\mathcal{M}^\epsilon$ where $0 \leq \epsilon \leq 1$. When $\epsilon = 0$, the system is completely compensated without any copy number dependence. On the contrary, when $\epsilon = 1$, transcription rate is linearly proportional to the copy number. The experimental data (Figure 15(A)) supports that ϵ is between 0.4 and 0.6. For stochastic simulations, we replace all the terms of $\mathcal{M}$ in Equation (63) with $\mathcal{M}^\epsilon$, and set $\epsilon = 0.5$.

CHAPTER V

THE FUTURE IS NOW: PREDICTING CELL FATE BY MEASURING TEMPORAL DYNAMICS OF GENE REGULATION[1]

5.1 Introduction

Cells often respond to variable environmental conditions by undergoing a sequence of intracellular changes that lead to a marked change in phenotypes, e.g. a change in cell fate. Understanding the basis for cell fate determination is a fundamental challenge from developmental biology [58] to medicine [197, 152]. Changes in cell fates are often governed by the dynamics of gene regulatory networks (GRNs) that are, in part, driven by environmental stimuli [194]. However, environmental stimuli do not always drive cell fates deterministically. Gene regulation is intrinsically stochastic, and predicting cell fate given a known stimulus may be difficult [53, 183]. Stochasticity in cell fate determination is exemplified by observations of multiple cell fates in isogenic populations responding to almost identical environmental conditions. For example, only a small fraction of *E. coli* cells treated with antibiotics develop bacterial persistence [15], and only some *B. subtilis* cells exhibit competence in response to the onset of stationary growth phase [113].

One approach to predicting cell fate is to monitor a cell's internal state in hopes of identifying intracellular epigenetic changes that develop into one of typical gene expression patterns within an isogenic population. Historically, a number of biomarkers

[1]This chapter is in preparation for publication [86].

have been developed as proxies for the intracellular states of cells, including electrical properties, metabolic states, and morphological changes [1, 70, 59, 134, 56]. More recently, a common method is to measure gene expression using either high-throughput or high-resolution approaches. High-throughput gene expression profiling reveals distinct patterns of expressed genes at the population level [104, 3]. In contrast, real-time imaging of gene expression has become a powerful tool in revealing temporal dynamics of gene regulation [151, 162, 22, 36, 198, 105]. The high-resolution approach has advanced rapidly due to technical advances in single-cell imaging, owing to fluorescent proteins, fluorophore, and quantum dots [31, 32, 200]. These advances have improved our understanding of dynamical systems that govern cell fate progression [180]. Moreover, temporal gene expression may be especially useful as a biomarker in cell fate prediction if a cell's internal state becomes increasingly correlated to the future phenotypic change.

Although studies of cell fates have been addressed before [198, 36], the results have tended to be system specific while many questions remain unanswered. For example, can we predict accurately a specific cell's fate from measurements of the temporal dynamics of gene regulation *well before* the phenotypic change? Moreover, is it possible to systematically predict fates of many isogenic cells subject to nearly identical environmental stimuli? To answer these questions, we present a quantitative framework that translates temporal dynamics of gene regulation into predictability of cell fate as a function of time. In principle, the framework can be applied to an arbitrary cell fate determination system. The framework is based on the notion that even if cells are seemingly unpredictable at the population level, different subpopulations may emerge that lead to particular cell fates [75]. To demonstrate this, we first analyze simple models of "gene regulation", which are a random walk model and a positive feedback loop with external noise. In doing so, we find that intrinsic bias in intracellular dynamics must dominate external noise to facilitate cell fate prediction in advance. We

then apply our framework to single cell experimental data on phage-λ-infected *E. coli* and compare the result with predictions from a mechanistic model of lysis-lysogeny decision switch [87]. We find that divergence of multiple gene expression patterns leads to enhanced predictability over time, albeit with a tradeoff: the number of cells that have not experienced a phenotypic change decreases with time. In other words, the number of possible predictions goes down over time while cell fate prediction for a given cell becomes increasingly more accurate. We discuss the optimal timing for cell fate prediction when there are benefits and penalties associated with correct and incorrect predictions, respectively.

5.2 Results

5.2.1 Stages of cell fate determination: "priming" by temporal dynamics

Cells undergo phenotypic changes in response to environmental stimuli. Often such phenotypic changes are mediated by sequences of dynamical changes driven by GRNs. GRNs may cause significant changes in gene expression at early time points. These changes prime cells for specific cell fates, and later dynamical changes would not alter the fates of these primed cells. In such cases, detection of the early dynamical changes can facilitate dramatic improvement in cell fate prediction. This is illustrated in Figure 17. Prior to environmental stimulation, cells have the same gene expression pattern, and shortly after the stimulation cells are in a similar state of gene expression (Stage I). At this stage, measurement of gene expression would not provide much information on the future fate of a given cell. However, individual cells follow different trajectories of gene expression over time due to stochasticity in gene regulation, resulting in emergence of subpopulations with distinct gene expression patterns (Stage II). At this stage, gene expression patterns of subpopulations are weakly correlated with eventual cell fates. Over time intrinsic biases of GRNs becomes increasingly strong, and prevents cells from switching between subpopulations once it dominates

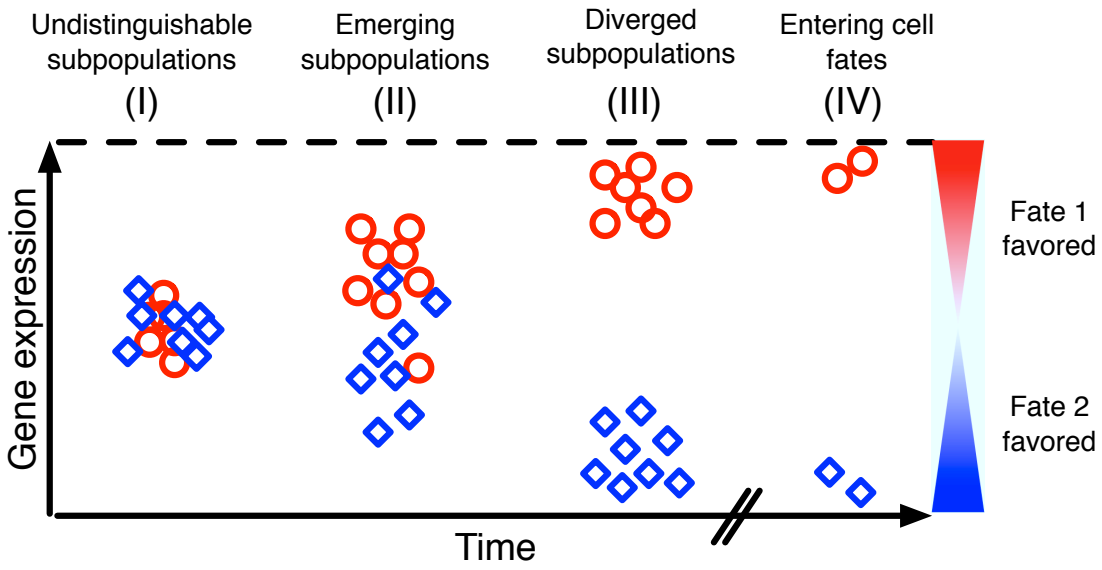

Figure 17: Stages of cell fate progression. A population of cells consists of two groups that ultimately reach either fate 1 (red o) or fate 2 (blue $\diamond$). Assume cells are initially prepared in the same condition. Near the initial condition cells have similar expression and the difference in gene expression between the two groups is undetectable (Stage I). These cells follow different trajectories over time due to stochastic gene regulation, leading to emergence of subpopulations that show distinctive gene expression patterns (Stage II). When the intrinsic bias of gene regulatory networks dominate stochasticity in gene expression, gene expression patterns of subpopulations further diverge from each other. Then gene expression patterns may become completely separated (Stage III), and cells in each subpopulation enter a specific cell fate (Stage IV). Note that cells may stay in Stage III for a significant amount of time before entering their final cell fates, as indicated by the broken axis, and cell fate prediction can be achieved well before the phenotypic change. The colored triangles represent probability of entering fate 1 or 2 at given gene expression.

stochasticity in gene regulation (Stage III). Thus, the future fate of any cell is already determined by gene expression at that time. Finally, cells start to enter their fates, and cell fate prediction is not applicable to those cells anymore (Stage IV). Here we focus on cases when the dynamics of subpopulations become deterministic well before they enter their final cell fates. In such cases, gene expression level becomes a useful indicator for cell fate prediction.

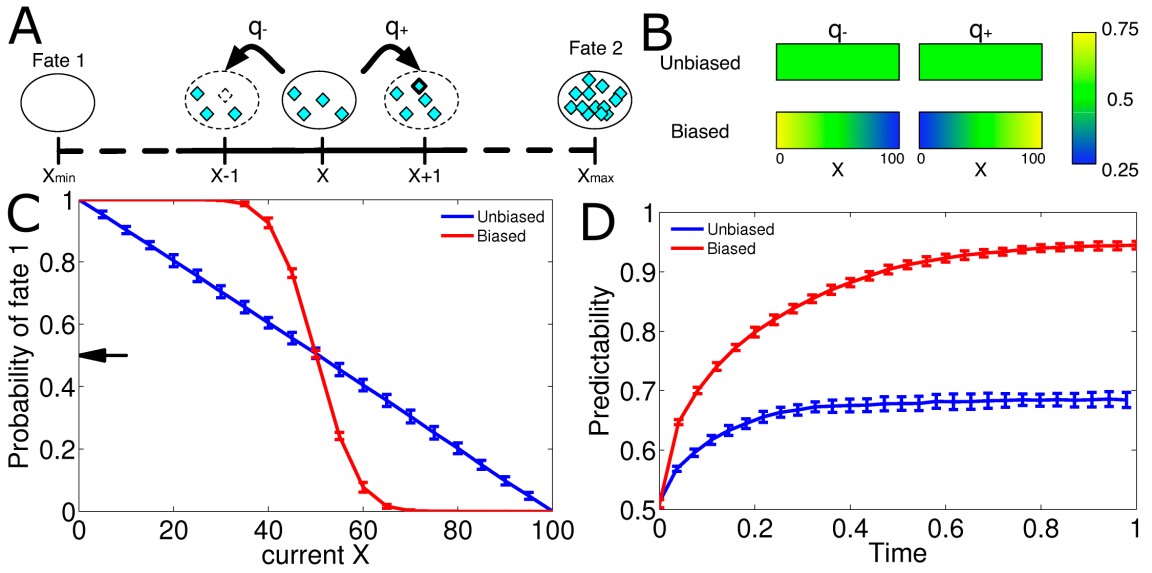

Figure 18: Effect of intrinsic bias on cell fate prediction. (A) A random walk model representing stochastic gene expression. The number of protein X can either increase or decrease by one molecule at a time and the transition probabilities from X to $X-1$ and $X+1$ are denoted as q_- and q_+, respectively. X_{min} and X_{max} are threshold values for entering fate 1 and 2, respectively. Once the gene expression reaches either threshold, the cell is assumed to irreversibly enter the corresponding fate. (B) Values of q_- and q_+ as functions of X. We consider two different conditions for random walk: 1) unbiased with fixed transition rates for all X, and 2) biased symmetrically toward both fates. (C) The probability of reaching fate 1 as a function of current gene expression. This is calculated by dividing the number of simulations that pass X and enter fate 1 by the number of all simulations that pass X. All the simulations start from $X = 50$ at $t = 0$, and we set $X_{min} = 0$ and $X_{max} = 100$. The black arrow represents the critical probability where a cell is equally likely to enter fate 1 and 2. At a given gene expression of a cell, the cell is predicted to enter fate 1 if the probability of fate 1 is greater than 0.5. Otherwise, we predict the cell would reach fate 2. Errorbars represent standard deviations of 10 independent sets of 1000 simulations. (D) Predictability as a function of time. Predictability is defined as the probability that the prediction based of temporal gene expression and cell's future fate are the same (see Section 5.4.1 for detail). In each case, the time is rescaled such that half the cells enter their fates by $t = 1$.

5.2.2 Intrinsic bias and predictability

An important assumption in our quantitative scheme is that gene expression changes are driven by nonlinearity and noise in GRNs. Dynamics of GRNs are often driven by deterministic skeletons (or intrinsic biases) that lead to biased changes in gene

expression as a function of the current cellular state [4]. Here we investigate how intrinsic biases in GRNs can affect predictability and render gene expression into an informative biomarker. We define predictability as the probability that our prediction of based on temporal dynamics of gene expression coincide with the eventual cell fate. To demonstrate the relationship between intrinsic dynamics biases and predictability, we begin with a simple random walk model with variable transition rates, which is a simplification of gene regulation (see Figure 18(A)). Dynamics of gene regulation is discrete due to discrete number of regulatory molecules in a cell. Each random walk with transition probabilities, q_- or q_+, corresponds to a decrease or increase in the number of molecules, respectively. Once gene expression passes a threshold level for fate 1 ($X_{min} = 0$) or fate 2 ($X_{max} = 100$), we assume that the cell immediately enters the corresponding fate, and the first passage to either threshold is equivalent to an irreversible commitment to a cell fate [148].

To demonstrate the effect of intrinsic bias on predictability, we analyze two different cases of random walks: (1) unbiased when the transition rates are the same for all X; and (2) biased symmetrically toward both fates. The transition probabilities under these conditions are shown in Figure 18(B). Our analyses of random walk under these conditions show that the probability of a cell fate at a particular value of X is strongly influenced by intrinsic bias as shown in Figure 18(C). In the unbiased case, probability of each fate scales with X due to a fixed transition rate for all X values. With symmetric biases toward both fates, a small deviation from the initial condition primes the cell for a specific fate. In this case the current concentration can serve as a good marker for cell's future fate. Based on the probability of each cell fate, we make a prediction for a cell given the current gene expression. If the probability of fate 1 is greater than 0.5, we predict the cell is likely to reach fate 1. Otherwise, the cell is expected to enter fate 2. Note that the gene expression which leads to predicting cell fate 1 or 2 is system dependent.

Changes in gene expression over time lead to changes of probabilities for each cell fate, and further change our cell fate predictions. In biased dynamical systems, gene expression is highly correlated with the probability for each cell fate, and this would lead to a rapidly increasing predictability with time. To demonstrate this, we calculate predictability, the probability of correctly predicting the fates of a cell with given gene expression, which is based on probabilities of each cell fate and resulting prediction. Here our system is totally symmetric, and each cell has equal chances of reaching fate 1 and 2. Therefore, we always expect half of them to enter fate 1 regardless of the time of measurements. However, if we look at gene expression of each cell, cells would start forming two subpopulations with distinct gene expressions patterns, and each subpopulation becomes increasingly correlated with a particular cell fate over time as illustrated in Figure 17. Our simulations in Figure 18(D) show that predictability indeed increases with time for both cases, but the symmetric bias leads to a significant improvement in predictability over the unbiased case. These results suggest a critical role of intrinsic bias in improving predictability.

5.2.3 Predictability and external noise

In addition to intrinsic biases, dynamics of gene regulation are governed by noise in a dynamical system. Noise might arise from two sources: 1) internal noise from small number of regulatory molecules and 2) external noise. In this section, we analyze the effect of external noise on predictability while incorporating the internal noise. External noise in gene expression may override any intrinsic bias and compromise predictability. To evaluate the effect of external noise on predictability, we analyze a positive feedback loop that exhibits bistability. Bistability has been shown to drive multiple cell fates in proliferation, differentiation, and lysis-lysogeny [179, 103, 184]. Upon environmental stimulation, the positive feedback loop dynamics can drive the cell to either low or high steady state gene expression, each of which corresponds

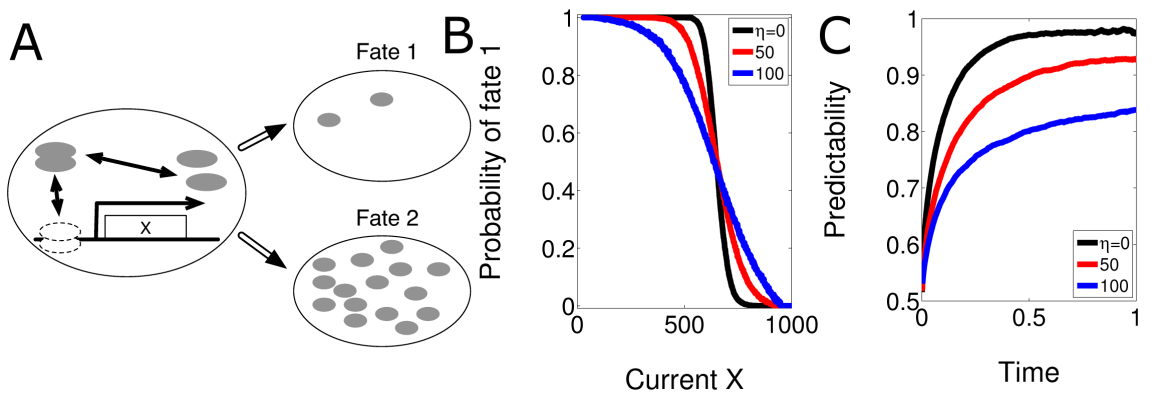

Figure 19: Effect of external noise on cell fate prediction. (A) Two cell fates driven by a positive feedback loop. A gene encoding protein X has a positive feedback on itself by self-promoting its activity via dimers. The positive feedback loop can generate low (fate 1) and high (fate 2) gene expression. Initially, we assume that cells are at an unstable state with intermediate gene expression. The dynamics of this positive feedback loop is shown in Section 5.4.3. (B) Probability of fate 1 with various external noise, η ($\eta = 0$ for black, $\eta = 50$ for red, and $\eta = 100$ for blue). To systematically investigate the effect of external noise on predictability, we added Gaussian noises with various amplitudes. η is the relative magnitude of external noise versus intrinsic fluctuations. Each curve represents the result from 10,000 independent stochastic simulations. Note that slope in the probability of fate 1 decreases with increasing η suggesting the external noise can offset intrinsic bias. (C) Predictability with various η. The predictability increases with time but decreases with increasing η ($\eta = 0$ for black, $\eta = 50$ for red, and $\eta = 100$ for blue).

to a cell fate as illustrated in Figure 19(A). The thresholds for entering fate 1 and 2 is defined as two steady state of gene expression (see Section 5.4.3 for details). To ensure equal bias toward both cell fates, we use appropriate initial conditions and input strength. We vary the level of external noise by changing η which is the relative magnitude of external and internal noise, and evaluate the effect of η on predictability via stochastic simulations (see Section 5.4.3 for further details on stochastic simulations).

Note that this system is biased, and small deviations from the initial condition would favor one cell fate over the other. Therefore, gene expression becomes increasingly correlated with future cell fates over time. With external noise, we expect predictability would increase slowly over time because gene expression will be weakly

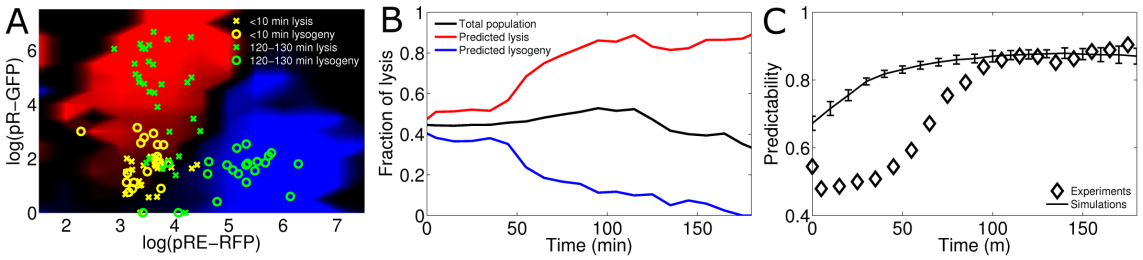

Figure 20: Analysis of single cell experimental data of phage-λ- infected *E. coli* **(courtesy of Zeng** *et. al.* **[201]).** (A) Temporal changes in gene expression. Color maps representing the probability of lysis (red) and lysogeny (blue) based on lytic reporter pRE-RFP and lysogenic reporter pR-GFP. This probability of lysis and lysogeny determines our prediction of a cell. If a cell exhibits gene expression in the blue region, we predict that the cell would enter lysogeny. A cell is expected to enter lysis with gene expression in the red region. For each cell, the final fate of cells is denoted by either x for lysis and o for lysogeny. The different colors of x and o represent different time points (yellow for <10 mins and green for 120-130 mins). Note that gene expression patterns are highly correlated with eventual cell fates after two hours. That is cells predicted with lysogeny and lysis will predominantly enter lysogeny and lysis, respectively, for most times. (B) Fraction of lysis for total population and subpopulations predicted with lysis and lysogeny. Note that the fraction of lysis within the total population stays almost constant over time, which suggest that cell fates are unpredictable at the total population level. However, if we look separately at each subpopulation based on temporal gene expression, each subpopulation shows increasing tendency of entering a particular cell fate over time. Thus, prediction of lysis and lysogeny becomes accurate as gene expression become increasingly correlated with cell fates (C) Predictability for experimental data (diamond) and simulations from a mechanistic model (solid line) of lysis-lysogeny decision switch [87]. The errorbar represents the standard deviation from 10 independent simulations with same number of cells from the experimental data (see Section 5.4.5 for details).

correlated to cell fates. Our analysis in Figure 19(B) demonstrates that the probability of reaching more dominant cell fate at any given X decreases with increasing η. Under an extreme level of η, noise may completely override the intrinsic bias, in which case the probability curve would be similar to the unbiased case in Figure 18(C). In addition, Figure 19(C) shows decreasing predictability with increasing η. Together, these results demonstrate how external noise can negatively impact cell fate prediction. Nonetheless, prediction of cell fate remains possible for a wide range of external noise levels.

5.2.4 Predictability for phage-λ-infected *E. coli*

Phage λ is one of the simplest organisms exhibit alternative cell fate determination within infected bacterial hosts [146, 97, 137]. Lysis occurs when infecting phages kill the host to release daughter phages. Phages can also integrate their genomic material into the host chromosome to transmit vertically with the host, a phenomenon known as lysogeny. Phage λ is an excellent model system to apply our quantitative framework for the following reasons. First, cell fates are determined within a few hours after infection and are clearly distinguishable. Second, the core GRN consists of various feedbacks that induce bistability (or at least transient bistability [87]) in the underlying dynamics. Third, fluorescence-tagged proteins that report the temporal dynamics of viral gene expression are readily available. Here we analyze single cell experimental data previously published by Zeng *et. al.* [201].

Our analyses of the experimental data show that lysis is strongly correlated with high expression of a lytic reporter pR-GFP and low expression of a lysogenic reporter pRE-RFP (red lytic regime) as shown in Figure 20(A). In contrast, lysogeny is strongly correlated with low pR-GFP and high pRE-RFP (blue lysogenic regime). We predict a cell would enter lysis or lysogeny depending on if the gene expression is in the lytic or lysogenic regime. Shortly after infection, most cells have low probability of entering both lysis and lysogeny as most cells have gene expression near the border between the lytic and lysogenic regime (see yellow + and o in Figure 20(A)). Over time, however, most cells that we predict to enter lysis or lysogeny actually enter lysis or lysogeny, and their gene expression moves deeper into either the lytic or lysogenic regime (green o and +), leading to divergence of the cell populations based on cell fate prediction by temporal gene expression. The divergence becomes increasingly clear with time as shown in Figure 20(B). While the fraction of cells that enter lysis remains nearly constant in the total cell population, the fraction of cells which later enter lysis with gene expression in the lytic regime increases over time. Moreover, the

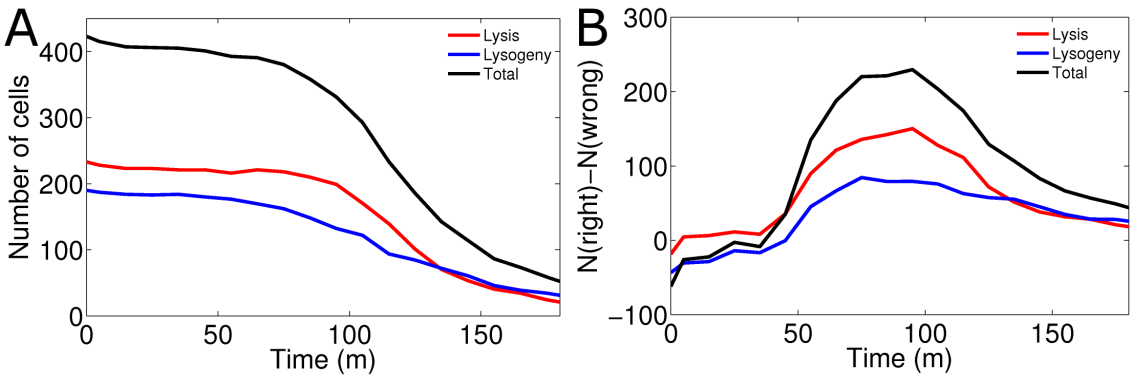

Figure 21: Tradeoff between predictability and its scope. (A) The number of cells that have not reached their final fates decreases as cells reach their final fates over time. (B) The difference between correct and incorrect predictions is a non-linear function with time. This represents the benefit and penalty of making correct and incorrect predictions, respectively. If ultimate fate of a cell coincides with the predicted cell fate by gene expression at a given time, the prediction is considered correct. Otherwise, the prediction is considered incorrect. Predictions are made only on the cells that have not reached their final fates.

fraction of cells entering lysis in the lysogenic regime decreases with time.

We have demonstrated that increasing divergence into subpopulations over time may improve predictability with our simple models (Figures 18–19). Our analysis of experimental data also confirms increasing predictability with increasing divergence over time, as shown in Figure 20(C). Remarkably, this trend can be recapitulated with our mechanistic model on lysis-lysogeny [87] (see Section 5.4.5 for detail). However, there is a large time delay in predictabilities from the experimental data and simulations of the mechanistic model. The time delay in the experimental predictability is most likely due to a delay in DNA injection [202, 135] and transcription anti-termination [146]. In the experimental data, there are some cells which do not develop into either lysis or lysogeny, and we refer them as nave cells. Nave cells are excluded in the calculation of predictability, and their predictability does not increase with time. This suggests that pRE-RFP and pR-GFP are not appropriate biomarkers for predicting naive cells.

5.2.5 Tradeoff between predictability and its scope

We have shown that intrinsic bias underlies an increasing correlation between the final fate of a cell and its gene expression over time, and predictability increases over time. However, increase in predictability does not guarantee better predictions for a population of cells. It may neglect an intuitive observation: an increasing number of cells will reach their final cell fates over time as shown in Figure 21(A). When we combine this observation with predictability reveals an important insight for cell fate prediction based on temporal dynamics of gene regulation. Because the predictability increases over time, we can make increasingly better prediction for a given cell. However, the number of cells to be predicted decreases as cells reach their eventual cell fates. This suggests a tradeoff between predictability and its scope. To demonstrate this tradeoff, we use the difference between correct and incorrect number of predictions as a possible optimization criterion for cell fate prediction (Figure 21(B)). A prediction of a cell fate is considered correct if the ultimate fate of a cell is same as the predicted fate by gene expression at a given time. Otherwise, a prediction is considered incorrect. Therefore, the difference between correct and incorrect number of predictions serves as a single metric that represents benefits (correct predictions) and costs (incorrect predictions) associated with cell fate prediction at a given time [127]. In this case, cell fate prediction is optimal when the difference is the greatest. Initially, the difference is negative due to more incorrect predictions, and this shows that cell fate prediction based on early gene expression can be costly rather than beneficial. With increasing predictability, cell fate prediction becomes increasingly beneficial, and becomes optimal at an intermediate time. This time point corresponds to a moderate predictability and a large number of remaining cells to be predicted. As more cells reach their final fates, the remaining cells become fewer over time. Note that the difference between correct and incorrect predictions is bounded by the total number of predictions, and this leads to a decreasing difference over time. These analyses

demonstrate a counter-intuitive observation that cell fate prediction is optimal with moderate predictability and a large number of cells to be predicted. This is because of strong intrinsic bias that primes cells towards their final fates well before they reach their fate. In other words, moderate predictability is sufficient to facilitate correct predictions of 'primed' cells toward cell fates. The tradeoff between predictability and its scope is a form of speed-accuracy tradeoff due to the discount in the total possible number of predictions over time [34].

5.3 Discussion

Focusing on the temporal dynamics of gene regulation, we have shown that divergence of subpopulations biased towards particular fates can facilitate cell fate prediction well before a marked change in phenotypes. In particular, we have developed a statistical framework that translates temporal dynamics of gene regulation into predictability as a function of time. Using simple models of gene regulation, we show that intrinsic biases can confer increasing predictability over time, but only if noise does not override intrinsic biases. Moreover, we apply our statistical model to experimental data of phage-infected bacteria and show that divergence of gene expression patterns driven by intrinsic bias indeed improves predictability over time. Remarkably, this can be recapitulated with our previously developed mechanistic model [87]. Despite increasing predictability over time, we have also shown that the scope of predictability is limited by the number of remaining cells as they enter their final cell fates. The increasing predictability and decreasing scope over time leads to a tradeoff that can be optimized by considering benefits and costs associated with cell fate prediction.

Our findings in this work suggest an optimal set of biomarkers and experimental strategies customized for a given system. This may be particularly important if phenotypic changes result from fast changing dynamics [65] driven by intrinsic bias and are independent of following slow dynamics (Figure 20). For such systems, only

the biomarkers customized to capture fast dynamics will prove useful. With these biomarkers it is critical that a high signal-to-noise ratio is maintained, as sufficiently strong noise may override the intrinsic bias that drives fast changing dynamics (Figure 19). Here we considered cell fate determination as a Markovian process determined only by the current gene expressions. However, history of cells can heavily influence their future cell fate [30, 23] because gene expression reported by biomarkers reveals only a tiny fraction of a cell's internal state. In such cases, prediction of the cell's future based only on its current state would be difficult and may depend on the entire history of gene expression. Finally, it has been shown that monitoring multiple genes does not always lead to enhanced assessment of a cell's state [191, 189]. This is due to overlapping dynamics of these genes, highlighting that a dynamical understanding of biomarkers is critical in designing appropriate biomarkers.

In our analysis, we assumed that the benefit of a correct prediction and the cost of an incorrect prediction are exactly opposite and the optimal tradeoff arises when the difference between correct and incorrect number of predictions is the greatest. However, depending on the context the benefits and costs may need to be weighted unequally based on the consequences of correct and incorrect predictions. For example, breast cancer is very common, and can be fatal if diagnosed in a late stage [120]. In such case, the cost of a wrong prediction (false negative) would outweigh the benefit of a correct prediction. In addition, benefits and costs associated with cell fate prediction can have broader implications than counting simply correct or incorrect predictions. Benefits may also include the gain of time interval between when a correct prediction is made and when the cell reaches its final fate, while costs may include experimental efforts and materials to assess cell states at the given time. These broader classes of costs and benefits will lead to different optimization results.

Then how can we find a system with potentially high predictive power? We have

demonstrated gain of predictive power over time through our models and experimental data. However, it may not be easy to find a good model system or biomarker for cell fate prediction. We expect temporal dynamics of gene expression may confer significant predictive power under two conditions. First, cell fates are mediated by the output of a coherent feed forward loop which has two inputs. Inputs may turn on significantly earlier that the output, hence may serve as a cue for output activation [116]. Second, when cell fates emerge as the result of a sequence of gene regulation [88, 196, 157], gene expression intermediate genes is the key to determine whether or not a cell would develop a specific fate.

Our statistical framework is based only on temporal dynamics of gene expression patterns over time in a cell population. In principle, our framework can be generalized to other biomarkers including morphological and metabolic states, given their states diverge and become increasingly deterministic over time. For example, there are several studies which try to combine cell culture data and tumor imaging to study the relationship between cell proliferation and morphological characteristics [158]. In addition, these biomarkers, combined with gene expression patterns, may synergistically enhance predictability.

5.4 Methods and Models

5.4.1 Calculation of predictability by temporal gene expression

Assume that there are N possible alternative phenotypes, and the dynamics of gene regulation are autonomous (not time-dependent). All the phenotypes are mutually exclusive. The temporal gene expression of a cell at a given time is represented as $\vec{X(t)} = (x_1(t), x_2(t), ..., x_d(t))$ where d is the dimension of gene expression states. Consider a collection of time course trajectories of M cells starting from a known distribution of initial conditions. The cell fate of j-th cell is denoted as D_j ($1 \geq D_j \geq N$). Assume that cells are identical and independent from one another (no cell-to-cell

interactions). We are interested in which fate will a cell enter given a cell's current gene expression. The predictability, $\mathcal{P}(t)$, is the probability that the prediction based on temporal gene expression coincides with the eventual cell fate as a function of time. $\mathcal{P}(t)$ is calculated as the following:

1. Sort the cells into d-dimensional bins. A cell is assigned to a bin if the cell have observed with gene expression corresponding to the bin at any time.

2. Calculate the number of cells for each bin.

3. Within a bin, count how many cell end up with i-th fate and calculate the fraction of i-th decision, $f_i(\vec{x})$, in the bin. $f_i(\vec{X})$ is the probability of i-th fate when a cell have gene expression of $\vec{x}$.

4. Make cell fate predictions based on temporal gene expression for a cell at a given time. For each cell at a given time, compare the probabilities of all fates and predict the most likely fate. A cell is predicted to enter i-th fate at time t if $max\{f_j(\vec{X}(t)\}_{j\in\{1,2,\ldots,N\}} = f_i(\vec{X}(t))$.

5. At a given time, make predictions for all cells based on gene expression at that time. Predictability, $\mathcal{P}(t)$, becomes the average of the probability of the most likely fates for all cells.

5.4.2 Random walk model

We consider a random walk process in gene expression with a fixed step size. The expression can change from X to either $X+1$ or $X-1$. Initial condition is always $X = 50$. When the particle reach $X = 0$, it commits to fate 1 while reaching $X = 100$ leads to fate 2. Once a cell reach either decision, that cell is removed from the population, and there is no further prediction. The transition probabilities are

denoted by q_- and q_+:

$$X \xrightarrow{q_+} X + 1\,, \tag{66}$$

$$X \xrightarrow{q_-} X - 1\,. \tag{67}$$

The values of q_- and q_+ is shown in Figure 18(B). Note that depending on q_- and q_+ the time scale of entering cell fates can be different, and it is shorter when the system is biased than when unbiased. To correct the difference in the time scale, we rescale the time so that half the cells enter their corresponding cell fate by $t = 1$ (as shown in Figure 18(D)).

5.4.3 Model of the positive feedback loop

To model a simple positive feedback loop, we use the Hill kinetics with the Hill coefficient of 2 as the following:

$$\frac{dX}{dt} = \frac{\alpha + \beta X^n}{K^n + X^n} - mX\,. \tag{68}$$

where X represents the protein number of a gene. α, β, and mX denote a basal synthesis rate constant, a rate constant for the synthesis of X driven by environmental stimulation and degradation constant. For stochastic simulations of the positive feedback loop, we adopt the chemical Langevin formulation [61] which allows for implementation of intrinsic (reaction-dependent) and external (reaction-independent) noise.

To ensure equal bias toward fate 1 and fate 2, we use appropriate initial condition (X0=650) and input strength (β=1.03). Other parameters used in stochastic simulations are as follow: α=0.02 and m=0. The threshold for fate 1 is defined as the minimum level of X at which the system is at the high state (X=950). Similarly, the threshold for fate 2 is defined as the maximum level of X at which the system is at the low state (X=80).

5.4.4 Single cell experiments of phage-λ-infected *E. coli*

Each *E. coli* cells was exposed to fluorescently labeled phage λ, and infected with from 0 to no more than 10 phages. Cells infected with no phages are excluded from the analysis. Phage absorption occurs at low temperature, and phage injection starts only when there is an abrupt increase of temperature. Thus, multiple coinfection occurs almost simultaneously. There are 523 experiments, and 845 trajectories in total. Extra trajectories are due to cell division during the experiment. After cell division event, the divided cells are discounted by the number of cell division in the lineage, so that the estimation of odds ratio is not dominated by rapidly dividing cells. Here we excluded naive cells which do not show any signs of phage infections from the analysis.

5.4.5 Mechanistic model of the lysis-lysogeny decision switch

We use a previously published model by Joh and Weitz [87] which consists of mRNAs and proteins of 4 genes, *cI, cro, cII and Q*, and an antisense mRNA, *antiQ* (see Figure 11 and Equation (63)). The parameters are transiently divergent (see Table 2), and the degree of partial gene dosage compensation is 0.5 (see Section 4.2.4). The host cell volume is various from $0.5\mu m^3$ to $2\mu m^3$. Each simulation has 523 independent simulations, and for each simulation the number of coinfecting phage is drawn from the distribution of coinfection for experimental data (see Table 6 in Appendix C).

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

There are decision switches in many biological systems. The initial motivation of this dissertation came from the experimental observation that even temperate phages can systematically alter their behaviors in response to external environmental conditions which in this case are multiple coinfections [101]. Phages do not have any sensory organs to interact with one another, and phage-phage interactions should happen within a host cell. In this study we proposed a deterministic and stochastic model to demonstrate switching from lysis to lysogeny with increasing number of coinfections. However, this switching affects not only the gene expression pattern but also the population dynamics of phages and hosts in natural environments. A population model including hosts and phages would be necessary to determine if the switching is evolutionarily optimized.

We also showed that copy-number-sensitivity of gene expression arises from multiple network motifs. Within a genome, different segments undergo different copy number changes during cell division, and the location of a motif within a genome determines the degree of copy number variation. Thus, copy-number-sensitive genes may be more likely to be far from the origin of replication, and this spatial distribution can be tested by comparing whole genome sequences and protein-protein interaction networks. Also, to test the hypothesis of the partial gene dosage compensation I propose an experimental study for phage-infected bacterial hosts.

6.1 Summary of major findings

6.1.1 Effect of copy number of genetic components on gene expression

Our quantitative model of the lysis-lysogeny decision switch is one of the first models which included the copy number of genetic components as an explicit parameter of the model. This feature allowed systematic study of gene expression as a function of copy number. Results from our model suggest that phages can alter their behaviors collectively in response to the number of coinfections. Interactions among phages occur via the transcriptional responses of individual phage genomes to shared gene products, in this case the regulatory proteins and mRNAs associated with the lysis-lysogeny decision switch. Each phage has the same viral genome, and shares its genetic circuits and gene products with one another. Strong nonlinearity within transcriptional feedback is the critical feature which facilitates bistability and drastic change in gene expression by small change in copy number. The range of viral concentration with bistability depends on all kinetic parameters in the quantitative model, which suggests that collective decision making in response to multiple coinfection is a tunable evolutionary feature of the lysis-lysogeny decision switch.

We also demonstrated that this copy-number-sensitivity of gene expression patterns occurs in other network motifs. In agreement with the lysis-lysogeny decision switch, sufficiently strong nonlinearity in the transcriptional feedback is the key feature that leads to bistability. In more complex GRNs, changing the copy number of a network motif can lead to qualitative change in gene expression patterns. The gene expression pattern from the repressilator shows that either stationary or oscillatory gene expression is possible depending on the copy number of the repressilator motifs. These results suggest that there may be class of genes and pathways which can readily develop novel gene expression patterns or new functions by duplication events.

6.1.2 Decision switches mediated by temporal dynamics of gene expression

We used a mechanistic model based on first passage processes of regulatory protein concentrations to determine the exact timing of entering different cell fates. Our model is the first stochastic model which incorporated the viral concentration into determining the fraction of lysogeny. In doing so, we quantified the stochastic response of phage λ to multiple coinfections, and our results recapture the increasing fraction of lysogeny with higher viral concentration. Previous experimental data suggest that at a given viral concentration there is a systematic variation in the fraction of lysogeny depending on the number of coinfections [201]. To explain this variation we proposed a hypothesis of partial gene dosage compensation as an alternative mechanism, and our hypothesis also demonstrate the systematic variation in response to multiple coinfections. Moreover, our hypothesis is experimentally falsifiable.

Recent studies showed that the current state of a cell can be an indicator for future cell fates [198, 36]. Here we use the temporal dynamics of gene expression as the indicator, and showed that predictability, the probability of making correct predictions, increases over time. This is due to emergence of qualitatively distinct subpopulations by temporal gene expression, and cells in each subpopulation preferentially enter one cell fate. Along with increase in predictability over time, cells enter their corresponding cell fate over time. Thus, over time prediction on a specific cell's fate become increasingly accurate, but the total number of possible predictions decreases. This suggest that there is a tradeoff of increasing predictability and its scope over time.

6.2 Study limitations

6.2.1 Effect of hosts on the lysis-lysogeny decision switch

Here we considered the lysis-lysogeny decision switch is mainly driven by gene regulation of phage λ. The only factor from hosts that can affect cell fates is the volume of host cells. However, there are other host factors which control the decision of lysis and lysogeny. CII is a protein that is unstable and critical to establishing lysogeny. There is a known protease HflA which cleaves CII into smaller fragments [94]. This suggests that the physiological states of hosts other than the cell size may greatly affect the bias between lysis and lysogeny. Moreover, phage λ have a protease inhibitor CIII to regulate Hf1A activity [98], and interplay between phages and host cells might be more complex than previously thought.

We assumed that the bacterial cytoplasm is well mixed at all times, but there are increasing number of evidences that bacterial cytoplasm is not well mixed [182, 156]. Thus, localization of viral proteins may play a critical role determining infected host cell's fate, and interaction among coinfecting phages may be limited due to localization of viral proteins. Moreover, phage λ is also known to preferentially inject viral DNA at the cellular poles and equator [51, 201], and replicate viral DNAs near the region of DNA injection. This suggests that there may be some host factors governing the decision of lysis and lysogeny locally, and the bias between lysis and lysogeny might be different between cellular equator and poles.

6.2.2 Defective phages

The single cell experiments by Zeng *et. al.* showed that a significant fraction (up to 20%) of singly infected cells do not exhibit any signs of phage infection [201]. Although host cells might play a role by dissecting foreign viral DNA through a mechanism like clustered regularly interspaced short palindromic repeats [16], phages are often defective. It is known that phages made by lysogens are more likely to be

defective [29] because viral DNA undergoes mutational decay during the lysogenic life cycle which can be orders of magnitudes longer than the generation time of bacterial hosts. This suggests that the number of viral genomes which successfully invaded a host cell might be lower than the number of coinfections. Also, it is not known if the fraction of phages that fail to infect a host is identical for each phage or if the fraction of failed infection depends on the number of coinfections.

6.2.3 Complexity of the lysis-lysogeny decision switch

In this dissertation, we only considered only a fraction of genetic components within the lysis-lysogeny decision switch for mathematical tractability. However, the entire circuit governing the decisions of lysis and lysogeny is much more complex with more than 10 viral genes [146]. A critical component that is not included in our models is the antiterminator, N, and antiterminator proteins control gene regulation by preventing transcription termination [62]. At sufficiently high N, genes downstream of transcription termination sites will be transcribed. Thus, antiterminators can change the topology of the GRN in a density-dependent manner. Also, in our models we did not consider genes downstream of Q and CI, and assumed that an infected cell enters lysis or lysogeny immediately after reaching the lytic or lysogenic threshold concentration. In reality, there can be a significant time delay between reaching either threshold and entering a cell fate due to the time to produce the genes downstream of Q and CI.

6.2.4 Dynamics thresholds for first passage processes

In Chapter 4, we assumed that the decision of lysis and lysogeny is mediated by strict first passage processes of regulatory proteins. Although first passage processes are useful to determine the timing of entering each cell fate, they oversimplify the lysis-lysogeny decision switch. Typically a cell fate is triggered by turning on downstream pathways with mRNA transcripts or proteins. When these reactions occur depends

on gene expression and binding affinities. Thus, a more realistic scenario would be entering cell fates with some probability at a given gene expression. Since it is also unlikely that an infinitesimally short pulse of sufficiently high gene expression would lead to a cell fate, the time scale of sufficiently high gene expression may be an important feature. First passage processes are known to be sensitive to noise [121], but dynamic thresholds might lead to robust mean first passage time.

6.3 Future directions

6.3.1 Ecological and evolutionary study of host-phage dynamics

In this dissertation, we showed the viral concentration within a host cell deterministically and stochastically biases the fraction of lysogeny. However, how the viral concentration affects the interplay between phage λ and its bacterial host is largely an open question. Also, it is not clear if utilizing two alternative cell fates of lysis and lysogeny is beneficial to phages. If we consider the dynamics of free living bacteria and phages, the number of coinfections depends on densities of phages and hosts. Only when the time scale between two successive phage infections are shorter than the time scale of the lysis-lysogeny decision switch, multiple coinfections are expected. Thus, ecological factors can drive the decision of lysis and lysogeny.

Moreover, the decision of lysis and lysogeny further changes the environmental conditions for hosts and phages. Lysis directly leads to increases in phage density and facilitates multiple coinfections. On the other hand, hosts with integrated viral DNA, which are called lysogens, are immune to further infections [146] and may serve as a sink for phages. To investigate the effect of lysis and lysogeny fully at the ecological scale, a two-species population dynamics model would be necessary, and there should be two mutually exclusive groups for the host population, which represent susceptible hosts and lysogens. The dynamics between susceptible hosts and lysogens should incorporate two decision switches: 1) the lysis-lysogeny decision switch and

2) the prophage induction switch when lysogens enter into the lytic pathway. Thus, such a model should incorporate the full cycle of phage infections.

Such a model might reveal a number of interesting aspects of phage-host dynamics. First, the model might lead to direct calculation of fitness for both lysis and lysogeny. Estimating fitness of lysogeny has been difficult in part because there are multiple life stages which are much longer than generation time of hosts. Fitness of lysogens might reveal environmental conditions where lysogeny is evolutionarily favorable. Second, the model can determine whether or not the phage λ optimizes their behaviors in the natural setting. Sensitivity analysis of model parameters would indicate key parameters which govern the host-phage dynamics. Lastly, investigating the viral concentration dependence in other temperate phages may confer information on their ecological conditions. There are many other temperate phages other than phage λ that undergo the decision of lysis and lysogeny, but we know little about their natural environmental conditions. For such a phage, measurement of the viral concentration dependence on lysogeny and comparing it with phage λ model might lead to inference on dynamical features of host-phage dynamics such as natural density of hosts and phages, and the rate of the prophage induction.

6.3.2 Spatial distribution of copy-number-sensitive genes within a genome

We demonstrate that the copy number of various network motifs is a key parameter governing gene expression patterns. This suggests that the gene expression may be more sensitive to copy number changes than previously thought. Also, we identified certain motifs can be highly susceptible to copy number variations, i.e., small change in copy number might lead to drastic change in gene expression pattern. In slowly growing cells, the copy number of any genes within a genome is expected to be one. However, polyploidy, which means a cell have multiple copies of the same chromosome, occurs in many organisms, and some bacteria have more than 10 copies of their

chromosome under fast growing conditions [133, 114]. Moreover, the copy number of a gene varies depending on the distance from the origin of replication and growth rate. A recent study showed that early replicating genome regions are more highly expressed due to higher copy number than other regions in the same genome [7]. This suggests that genes near the origin of replication undergo greater copy number variation due to changes in growth rate than genes far from the origin. We expect that the gene expression pattern within a cell should be robust to copy number variation to maintain homeostasis. Thus, copy-number-sensitive genes might not be located equally across the genome. If there are copy-number-sensitivity dependent patterns within a genome, such patterns can be tested by combining known databases on protein-protein interaction networks and whole genome sequencing. First, by analyses of protein-protein interaction networks, we can identify lists of proteins of a given species which are potentially sensitive to copy number variations. Then the location of such proteins relative to the origin of replication can be mapped within a whole genome. If our hypothesis is true, copy-number-sensitive genes might be far from the origin of replication while early replicating genes would be insensitive and robust to copy number changes.

6.3.3 Testing the hypothesis of partial gene dosage compensation

Phage coinfections inhibit host cell growth [201], but the exact mechanism of growth inhibition is unknown. Here we proposed a hypothesis that partial gene dosage compensation occurs in infected host cells with compensated transcriptional activity as a potential mechanism. However, it is not clear if partial gene dosage compensation is limited only to viral genes. Maybe partial gene dosage compensation is a general mechanism within a host cell which applies to host chromosome, viral genomes and even plasmids. To test if partial compensation acts at the level of transcription regulation and the extent of compensation, we propose the following experiments. The first

experiment would be to fuse a constitutive fluorescent reporter into the viral genome of phage λ and measure the reporter expression level at various viral concentrations. This would serve as a direct test for compensation in viral genes. A reporter in the host chromosome which is regulated by viral proteins can serve as an alternative indirect test for compensation of viral genes because the level of compensation might differ in the viral genome and the host chromosome. To check compensation within the host genome, a constitutive reporter integrated into the host chromosome would be sufficient. All the proposed experiments are designed to check if the partial gene dosage compensation is limited only to the viral genome or applied even to the host chromosome. We can distinguish if the compensation is at the level of transcription or translation by applying RNA-seq to each proposed system [130].

6.4 Conclusions

We used simple quantitative models to investigate various aspects of the lysis-lysogeny decision switch. Modeling the decision switch based on gene expression indicates a possible mechanisms for multiple alternative behaviors. However, a multi-scale and multi-species model would be necessary to fully assess consequences of this decision switch under natural conditions. At least, we showed that even very simple GRNs can lead to multiple gene expression pattern and even respond to different environmental inputs. Due to the simplicity of our models, we expect they can be easily modified and applied to other decision switches.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

A.1 Positive Feedback

The dynamics of the positive feedback motif described in the main text are given by the following system of differential equations:

$$\dot{x} = 2\kappa_- y - 2\kappa_+ x^2 + \sigma m - \gamma_p x \tag{69}$$

$$\dot{y} = -\kappa_- y + \kappa_+ x^2 + k_- d_1 - k_+ d_0 y \tag{70}$$

$$\dot{d_0} = k_- d_1 - k_+ d_0 y \tag{71}$$

$$\dot{d_1} = k_+ d_0 y - k_- d_1, \tag{72}$$

$$\dot{m} = \alpha d_0 + \beta d_1 - \gamma_m m, \tag{73}$$

where x and y are the concentrations of protein monomers and dimers, respectively, d_0 and d_1 are the concentrations of unoccupied and occupied promoters, respectively, and m is the mRNA concentration. Parameters $\kappa_\pm$ denote the dimerization and de-dimerization rates, $k_\pm$ are the binding and dissociation rates of the dimers to the promoter site, σ is the translation rate, γ_m is the degradation rate of mRNA, γ_p is the degradation rate of monomers, α is the basal transcription rate, and β is the activated transcription rate. We ignore the degradation of dimers for the sake of analytic tractability.

To understand how the steady state behavior of the above system depends on copy number we shall derive a relation between $\mathcal{N}$ and the steady state monomer concentration. First, we use a quasi-steady state approximation to obtain a protein only analogue of the system. Note that translation of proteins is much slower than dimerization and binding. It is also slower than mRNA transcription. Therefore, we

may simplify the full system by assuming that the fast variables are always at their steady states. Setting the corresponding derivatives (that is, derivatives of all the variables except x) to zero we obtain the following set of equations:

$$\begin{aligned} 0 &= -\kappa_- y + \kappa_+ x^2 + k_- d_1 - k_+ d_0 y && (74)\\ 0 &= k_- d_1 - k_+ d_0 y && (75)\\ 0 &= k_+ d_0 y - k_- d_1, && (76)\\ 0 &= \alpha d_0 + \beta d_1 - \gamma_m m. && (77) \end{aligned}$$

Solving this system we find that

$$\begin{aligned} y &= c_p x^2, && (78)\\ d_1 &= d\frac{c_p c_d x^2}{1 + c_p c_d x^2}, && (79)\\ d_0 &= \frac{d}{1 + c_p c_d x^2}, && (80)\\ m &= \frac{d}{\gamma_m}\frac{\alpha + \beta c_p c_d x^2}{1 + c_p c_d x^2} && (81) \end{aligned}$$

where $c_p = \kappa_+/\kappa_-$, $c_d = k_+/k_-$, and the total concentration of promoter sites is $d = d_0 + d_1$. Substituting these expressions into Equation (69) and recalling that $d = \mathcal{N}C$ we obtain the following protein only analogue of the original system:

$$\dot{x} = \frac{\mathcal{N}C\sigma}{\gamma_m}\frac{\alpha + \beta c_p c_d x^2}{1 + c_p c_d x^2} - \gamma_p x. \quad (82)$$

Here, we do not make use of the dynamical prefactors introduced in [52, 20] since we focus solely on the steady state behavior of the system. Notice that steady states of this equation completely determine steady states of the original model via Equations (78–81). Therefore, it is enough for us to analyze the protein only analogue.

It is convenient to rescale the above equation. Setting $u = \sqrt{c_p c_d}x$, $\hat{\alpha} = C\alpha\sigma\sqrt{c_p c_d}/\gamma_m\gamma_p$, $\hat{\beta} = C\beta\sigma\sqrt{c_p c_d}/\gamma_m\gamma_p$, $\tau = \gamma_p t$, and denoting derivative with respect to τ by a prime we get

$$u' = \mathcal{N}\frac{\hat{\alpha} + \hat{\beta}u^2}{1 + u^2} - u, \quad (83)$$

The relation between $\mathcal{N}$ and the (rescaled) steady state monomer concentration, which we denote by $\bar{u}$, is obtained by setting the left hand side of Equation (83) to zero. We find that

$$\mathcal{N} = \frac{\bar{u}(1+\bar{u}^2)}{\hat{\alpha}+\hat{\beta}\bar{u}^2}. \tag{84}$$

Real non-negative solutions to the above equation are biologically relevant steady states of Equation (83). Regarding $\mathcal{N}$ as a function of $\bar{u}$, such solutions are precisely the intersections of horizontal lines $\mathcal{N} = const$ with the graph of the function. Notice also that Equation (84) is equivalent to a cubic equation and therefore can have one or three real non-negative roots, depending on the parameters. Thus, we can expect that changing copy number, $\mathcal{N}$, may lead to a transition from one to three (or vice-verse) steady states. Such a transition occurs through a saddle-node bifurcation, and the values of $\mathcal{N}$ when it happens can be found as solutions to the equation $\partial\mathcal{N}/\partial\bar{u} = 0$, that is

$$\frac{\hat{\beta}\bar{u}^4 - (\hat{\beta}-3\hat{\alpha})\bar{u}^2 + \hat{\alpha}}{(\hat{\alpha}+\hat{\beta}\bar{u}^2)^2} = 0 \tag{85}$$

Since the denominator is always greater than zero, we get the following bi-quadratic equation:

$$\bar{u}^4 - (1-3a)\bar{u}^2 + a = 0, \tag{86}$$

where $a = \hat{\alpha}/\hat{\beta} = \alpha/\beta$. Solving for $\bar{u}^2$ we find

$$\bar{u}^2 = \frac{1-3a \pm \sqrt{(1-3a)^2 - 4a}}{2}.$$

Since we are interested in simple real roots we need the right hand side to be real and positive. Therefore, we need

$$1-3a > 0, \tag{87}$$

$$(1-3a)^2 - 4a > 0. \tag{88}$$

Remembering that $a > 0$, the solution to these inequalities is given by $a < 1/9$, or equivalently $\beta > 9\alpha$. Thus, if this condition is satisfied then there will be two

saddle-node bifurcations at

$$\mathcal{N}_{\pm} = \mathcal{N}(\bar{u}_{\pm}) = \frac{\bar{u}_{\pm}(1+\bar{u}_{\pm}^2)}{\hat{\alpha}+\hat{\beta}\bar{u}_{\pm}^2},$$

where

$$\bar{u}_{\pm} = \sqrt{\frac{\beta-3\alpha}{2\beta} \pm \sqrt{\left(\frac{\beta-3\alpha}{2\beta}\right)^2 - \frac{\alpha}{\beta}}}.$$

Equation (84) has only one real positive root for very small or very large $\mathcal{N}$. Also, the derivative $\partial\mathcal{N}/\partial\bar{u}$ is positive for $\bar{u} \in (0, \bar{u}_-) \cup (\bar{u}_+, \infty)$ and negative for $\bar{u} \in (\bar{u}_-, \bar{u}_+)$. Thus, the system has a single steady state, which is a stable node, for $\mathcal{N} \in (0, \mathcal{N}_-) \cup (\mathcal{N}_+, \infty)$ and three steady states, two stable nodes and a saddle, for $\mathcal{N} \in (\mathcal{N}_-, \mathcal{N}_+)$.

A.2 Bistable Feedback

The mathematical model of the bistable feedback motif is given by the following system:

$$\dot{x}_i = 2\kappa_- y_i - 2\kappa_+ x_i^2 + \sigma m_i - \gamma_i x_i \tag{89}$$

$$\dot{y}_i = -\kappa_- y_i + \kappa_+ x_i^2 + k_- d_i - k_+ d_0 y_i \tag{90}$$

$$\dot{d}_0 = k_-(d_1+d_2) - k_+ d_0(y_1+y_2) \tag{91}$$

$$\dot{d}_i = k_+ d_0 y_i - k_- d_i, \tag{92}$$

$$\dot{m}_i = \alpha_i d_0 + \beta_i d_i - \gamma_m m_i, \tag{93}$$

$$\tag{94}$$

where $i = 1, 2$, x_i and y_i are the concentrations of monomers and dimers, respectively, d_0 is the concentration of unoccupied promoters, d_i are the concentrations of promoters occupied by the first and the second proteins, and m_i are the mRNA concentrations. Parameters α_i and β_i, are the basal and enhanced/suppressed transcription rates and γ_i are the monomer degradation rates. The rest of the parameters have the same meaning as in Appendix A.1.

Following the steps described in Appendix A.1, we now derive a relation between the steady state monomer concentrations and copy number. First, we obtain a protein only analogue of System of Equations (89–93) by setting derivatives of all the variables except x_1 and x_2 to zero. Solving the resulting system of algebraic equations we find

$$y_i = c_p x_i^2, \tag{95}$$

$$d_0 = \frac{d}{1 + c_p c_d (x_1^2 + x_2^2)}, \tag{96}$$

$$d_i = d \frac{c_p c_d x_i^2}{1 + c_p c_d (x_1^2 + x_2^2)}, \tag{97}$$

$$m_i = \frac{d}{\gamma_m} \frac{\alpha_i + \beta_i c_p c_d x_i^2}{1 + c_p c_d (x_1^2 + x_2^2)}, \tag{98}$$

where $c_p = \kappa_+/\kappa_-$, $c_d = k_+/k_-$. The protein only analogue is obtained by substituting these expressions into Equation (89):

$$\dot{x}_i = \frac{\mathcal{N} C \sigma}{\gamma_m} \frac{\alpha_i + \beta_i c_p c_d x_i^2}{1 + c_p c_d (x_1^2 + x_2^2)} - \gamma_i x_i, \tag{99}$$

where we also used the fact that $d = \mathcal{N}C$. Steady states of this simplified system, which we denote by $\bar{x}_i$, completely determine steady states of the original system via Equations (95–98) and satisfy the following equations:

$$0 = \frac{\mathcal{N} C \sigma}{\gamma_m} \frac{\alpha_i + \beta_i c_p c_d \bar{x}_i^2}{1 + c_p c_d (\bar{x}_1^2 + \bar{x}_2^2)} - \gamma_i \bar{x}_i. \tag{100}$$

Setting $\bar{u}_i = \sqrt{c_p c_d}\bar{x}_i$, $\hat{\alpha}_i = C\alpha_i \sigma \sqrt{c_p c_d}/\gamma_m \gamma_i$, and $\hat{\beta}_i = C\beta_i \sigma \sqrt{c_p c_d}/\gamma_m \gamma_i$ we obtain a rescaled version of the above equations:

$$0 = \mathcal{N} \frac{\hat{\alpha}_i + \hat{\beta}_i \bar{u}_i^2}{1 + \bar{u}_1^2 + \bar{u}_2^2} - \bar{u}_i. \tag{101}$$

This system is equivalent to

$$\frac{\mathcal{N}}{1 + \bar{u}_1^2 + \bar{u}_2^2} = \frac{\bar{u}_i}{\hat{\alpha}_i + \hat{\beta}_i \bar{u}_i^2}, \tag{102}$$

which implies

$$\mathcal{N} = \frac{\bar{u}_1 (1 + \bar{u}_1^2)}{\hat{\alpha}_1 + \hat{\beta}_1 \bar{u}_1^2} + \frac{\bar{u}_2^3}{\hat{\alpha}_2 + \hat{\beta}_2 \bar{u}_2^2}, \tag{103}$$

$$0 = \frac{\bar{u}_1}{\hat{\alpha}_1 + \hat{\beta}_1 \bar{u}_1^2} - \frac{\bar{u}_2}{\hat{\alpha}_2 + \hat{\beta}_2 \bar{u}_2^2}. \tag{104}$$

After a detailed analysis of the above relation [123] we find that if

$$\beta_1 > 9\alpha_1 \text{ and } \gamma_2^2\beta_1\alpha_1 > \gamma_1^2\beta_2\alpha_2$$

then the system undergoes at least two consecutive saddle-node bifurcations as we change copy number. Moreover, if

$$\gamma_2^2\beta_1\alpha_1 \gg \gamma_1\beta_2,$$

then *only* two bifurcations can happen for *reasonably small values* of $\mathcal{N}$.

A.3 Toggle Switch

We model the dynamics of the toggle switch motif with the following system of differential equations:

$$\dot{x_i} = 2\kappa_- y_i - 2\kappa_+ x_i^2 + \sigma m_i - \gamma_i x_i \quad (105)$$

$$\dot{y_i} = -\kappa_- y_i + \kappa_+ x_i^2 + k_- d_{i+1} - k_+ d_{0,i+1} y_i \quad (106)$$

$$\dot{d}_{0,i} = k_- d_i - k_+ d_{0,i} y_{i+1} \quad (107)$$

$$\dot{d_i} = -k_- d_i + k_+ d_{0,i} y_{i+1} \quad (108)$$

$$\dot{m}_i = \alpha_i d_{0,i} + \beta_i d_i - \gamma_m m_i, \quad (109)$$

where $i = 1, 2$, and to simplify the index notation we assume that $2+1 = 1$, that is the index wraps back to 1 once it becomes greater than 2. The variables and parameters have the same meaning as in Appendix A.2, except that $d_{0,i}$ and d_i denote free and occupied promoters of the i-th gene.

A protein only analogue of Equations (105–109) is obtained by setting derivatives

of all the variables except x_1 and x_2 to zero. We find that

$$y_i = c_p x_i^2, \tag{110}$$

$$d_{0,i} = \frac{d}{1 + c_p c_d x_{i+1}^2}, \tag{111}$$

$$d_i = d\frac{c_p c_d x_{i+1}^2}{1 + c_p c_d x_{i+1}^2}, \tag{112}$$

$$m_i = \frac{d}{\gamma_m}\frac{\alpha_i + \beta_i c_p c_d x_{i+1}^2}{1 + c_p c_d x_{i+1}^2}, \tag{113}$$

where $c_p = \kappa_+/\kappa_-$, $c_d = k_+/k_-$, and substituting these expressions into Equations (105–109) we obtain the following protein only model:

$$\dot{x}_i = \frac{\mathcal{N} C\sigma}{\gamma_m}\frac{\alpha_i + \beta_i c_p c_d x_{i+1}^2}{1 + c_p c_d x_{i+1}^2} - \gamma_i x_i. \tag{114}$$

To find a relation between copy number, $\mathcal{N}$, and the steady state monomer concentrations we set the right hand side of the above equations to zero. After some algebra we get the following system:

$$\mathcal{N} = \frac{\bar{u}_1(1 + \bar{u}_2^2)}{\hat{\alpha}_1 + \hat{\beta}_1 \bar{u}_2^2} \tag{115}$$

$$0 = \frac{\bar{u}_1(\hat{\alpha}_2 + \hat{\beta}_2 \bar{u}_1^2)}{1 + \bar{u}_1^2} - \frac{\bar{u}_2(\hat{\alpha}_1 + \hat{\beta}_1 \bar{u}_2^2)}{1 + \bar{u}_2^2}, \tag{116}$$

where we used the following rescaling: $\bar{u}_i = \sqrt{c_p c_d}\bar{x}_i$, $\hat{\alpha}_i = C\alpha_i\sigma\sqrt{c_p c_d}/\gamma_m\gamma_i$, and $\hat{\beta}_i = C\beta_i\sigma\sqrt{c_p c_d}/\gamma_m\gamma_i$.

A detailed analysis of the above relation [123] reveals that if the following conditions hold:

$$\alpha_1 > 9\beta_1,\ \alpha_2 > 9\beta_2,$$

and

$$v\left(U_+\left(\frac{\alpha_1}{\beta_1}\right), \frac{\alpha_1}{\beta_1}\right) < \frac{\beta_2}{\beta_1} v\left(U_-\left(\frac{\alpha_2}{\beta_2}\right), \frac{\alpha_2}{\beta_2}\right) < v\left(U_-\left(\frac{\alpha_1}{\beta_1}\right), \frac{\alpha_1}{\beta_1}\right),$$

where

$$U_\pm(a) = \frac{1}{\sqrt{2}}\sqrt{a - 3 \pm \sqrt{(a-3)^2 - 4a}},$$

and

$$v(x,a) = \frac{x(a+x^2)}{1+x^2},$$

then the system undergoes two consecutive saddle-node bifurcations as we change copy number.

A.4 Repressilator

We use the following standard mathematical model of the repressilator [20]:

$$\begin{aligned}
\dot{x}_i &= 2\kappa_- y_i - 2\kappa_+ x_i^2 + \sigma m_i - \gamma_p x_i && (117)\\
\dot{y}_i &= -\kappa_- y_i + \kappa_+ x_i^2 + k_- d_{i+1} - k_+ d_{0,i+1} y_i && (118)\\
\dot{d}_{0,i} &= k_- d_i - k_+ d_{0,i} y_{i+2} && (119)\\
\dot{d}_i &= -k_- d_i + k_+ d_{0,i} y_{i+2} && (120)\\
\dot{m}_i &= \alpha d_{0,i} - \gamma_m m_i, && (121)
\end{aligned}$$

where α and γ_p denote transcription and monomer degradation rates, which are the same of all three genes, and the rest of the variables and parameters have the same meaning as in Appendix A.2. We start our analysis by obtaining a quasi-steady state approximation of Equations (117–121). Since variables $y_i, d_{0,i}$, and d_i change much faster than x_i and m_i, we set their derivatives to zero. Solving the resulting system of algebraic equations we find

$$\begin{aligned}
y_i &= c_p x_i^2,\\
d_{0,i} &= \frac{d}{1+c_p c_d x_{i+2}^2},\\
d_i &= \frac{d c_p c_d x_{i+2}^2}{1+c_p c_d x_{i+2}^2},
\end{aligned}$$

where $c_p = \kappa_+/\kappa_-$, $c_d = k_+/k_-$. Substituting these expressions into Equation (117) and Equation (121) and recalling that $d = \mathcal{N}C$ we obtain the following QSSA:

$$\begin{aligned}
\dot{x}_i &= \sigma m_i - \gamma_p x_i, && (122)\\
\dot{m}_i &= \mathcal{N}\frac{C\alpha}{1+c_p c_d x_{i+2}^2} - \gamma_m m_i. && (123)
\end{aligned}$$

It is convenient to rescale this system by setting $u_i = \sqrt{c_p c_d} x_i$, $n_i = \sigma\sqrt{c_p c_d}/\gamma_p m_i$, $\hat{\gamma} = \frac{\gamma_p}{\gamma_m}$, $\hat{\alpha} = C\alpha\sigma\sqrt{c_p c_d}/\gamma_m\gamma_p$, and $\tau = \gamma_m t$. Denoting derivatives with respect to τ by a prime, we obtain

$$u_i' = -\hat{\gamma}(u_i - n_i), \tag{124}$$

$$n_i' = \frac{\mathcal{N}\hat{\alpha}}{1 + u_{i+2}^2} - n_i. \tag{125}$$

This system can be shown to have a single symmetric steady state given by $\bar{u}_i = \bar{n}_i = \bar{u}$, where $\bar{u}$ is the unique solution of the following cubic equation:

$$\bar{u} + \bar{u}^3 = \mathcal{N}\hat{\alpha}. \tag{126}$$

We shall now investigate how stability of this steady state depends on the copy number. In particular, we shall see whether changing $\mathcal{N}$ can cause a Hopf bifurcation.

Let us rewrite System of Equations (124–125) in the vector form:

$$\begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \dot{u}_3 \\ \dot{n}_1 \\ \dot{n}_2 \\ \dot{n}_3 \end{pmatrix} = \begin{pmatrix} -\hat{\gamma}(u_1 - n_1) \\ -\hat{\gamma}(u_2 - n_2) \\ -\hat{\gamma}(u_3 - n_3) \\ \frac{\mathcal{N}\hat{\alpha}}{1+u_3^2} - n_1 \\ \frac{\mathcal{N}\hat{\alpha}}{1+u_1^2} - n_2 \\ \frac{\mathcal{N}\hat{\alpha}}{1+u_2^2} - n_3 \end{pmatrix}. \tag{127}$$

A simple computation shows that the Jacobian of the right hand side at the steady state can be written as a block matrix:

$$J = \begin{pmatrix} -\hat{\gamma} I_3 & \hat{\gamma} I_3 \\ -2f(\bar{u})A_3 & -I_3 \end{pmatrix},$$

where $f(\bar{u}) = \bar{u}^2/1 + \bar{u}^2$, I_3 is a 3×3 identity matrix, and

$$A_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

A Hopf bifurcation happens when a pair of conjugate eigenvalues of the Jacobian cross the imaginary axis. It can be shown that that the characteristic polynomial of the Jacobian is given by

$$\det(J - \lambda I_6) = (\hat{\gamma} + \lambda)^3(1 + \lambda)^3 + 8\hat{\gamma}^3 f^3(\bar{u}).$$

It is now easy to find the eigenvalues. In particular, the real part of the conjugate eigenvalues that can cross the imaginary axis is given by:

$$-\frac{\hat{\gamma} + 1}{2} + \frac{1}{2\sqrt{2}}\sqrt{\sqrt{a^2 + b^2} + a},$$

where

$$a = (1 - \hat{\gamma})^2 + 4\hat{\gamma} f(\bar{u}), \quad b = 4\sqrt{3}\hat{\gamma} f(\bar{u}).$$

To find when the corresponding eigenvalues cross the imaginary axis we equate the above expression to zero. After some simplification we obtain the following equation:

$$3\hat{\gamma} f^2(\bar{u}) + (1 + \hat{\gamma})^2 f(\bar{u}) - (1 + \hat{\gamma})^2 = 0.$$

Solving it with respect to $f(\bar{u})$ and taking into account that $f(\bar{u}) = \bar{u}^2/(1 + \bar{u}^2) > 0$ we obtain

$$f(\bar{u}) = g(\hat{\gamma}), \tag{128}$$

where

$$g(\hat{\gamma}) = \frac{2(1 + \hat{\gamma})}{1 + \hat{\gamma} + \sqrt{(1 + \hat{\gamma})^2 + 12\hat{\gamma}}}.$$

A simple analysis of the functions $f(\bar{u})$ and $g(\hat{\gamma})$ shows that Equation (128) has a unique solution for each fixed $\hat{\gamma}$. Employing Equation (126), the copy number at which the Hopf bifurcation occurs is given by:

$$\mathcal{N} = \frac{\sqrt{g(\hat{\gamma})}}{\hat{\alpha}(1 - g(\hat{\gamma}))^{\frac{3}{2}}}. \tag{129}$$

A.5 Kinetic Values

Our parameter estimates are approximate. They are in range with experimental measurements and typical values for dimerization, binding, transcription, translation and degradation in bacteria and viruses [4]. The approximate kinetic values common for the four motifs are as follows: $\kappa_+/\kappa_- = 10^7\ \mathrm{M}^{-1}$, $k_+/k_- = 10^7\ \mathrm{M}^{-1}$, $\sigma = 0.5\mathrm{min}^{-1}$, $C = 10^{-9}\mathrm{M}$, $\gamma_m = 0.1\mathrm{min}^{-1}$. Kinetic values specific for each of the motifs are:

Positive feedback: $\alpha = 0.025\mathrm{min}^{-1}$, $\beta = 1.7\mathrm{min}^{-1}$, $\gamma_p = 0.05\mathrm{min}^{-1}$.

Bistable feedback: $\alpha_1 = 0.1\mathrm{min}^{-1}$, $\beta_1 = 2.4\mathrm{min}^{-1}$, $\gamma_1 = 0.08\mathrm{min}^{-1}$, $\alpha_2 = 0.8\mathrm{min}^{-1}$, $\beta_2 = 0.01\mathrm{min}^{-1}$, $\gamma_2 = 0.06\mathrm{min}^{-1}$.

Toggle switch: $\alpha_1 = 10.12\mathrm{min}^{-1}$, $\beta_1 = 0.22\mathrm{min}^{-1}$, $\gamma_1 = 0.08\mathrm{min}^{-1}$, $\alpha_2 = 5.2\mathrm{min}^{-1}$, $\beta_2 = 0.52\mathrm{min}^{-1}$, $\gamma_2 = 0.06\mathrm{min}^{-1}$.

Repressilator: $\alpha = 2\mathrm{min}^{-1}$, $\gamma_p = 0.08\mathrm{min}^{-1}$.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

B.1 Probability of an active promoter

We assume that promoters are in thermodynamics equilibrium with transcription factors, and given concentrations of transcription factors the probability of a specific configuration is constant. Table 3 and Table 4 list all the possible configurations of promoters. The probability of i-th configuration is given as

$$f_i = \frac{exp(-\Delta G_i/RT)[CI_2]^{k_i}[RNAP]^{j_i}[CRO_2]^{l_i}}{\sum_i exp(-\Delta G_i/RT)[CI_2]^{k_i}[RNAP]^{j_i}[CRO_2]^{l_i}} \; for \; P_R/P_{RM} \qquad (130)$$

$$g_i = \frac{exp(-\Delta G_i/RT)[CII_4]^{m_i}[RNAP]^{j_i}}{\sum_i exp(-\Delta G_i/RT)[CII_4]^{m_i}[RNAP]^{j_i}} \; for \; P_{RE}\,, \qquad (131)$$

where G_i is free energy. f_i and g_i is the probability of i-th configuration for P_R/P_{RM} and P_{RE}, respectively. j_i, k_i, l_i and m_i represent the numbers of RNAP, CI, CRO and CII bound to the promoter. Note that a single RNAP occupies both O_{R1} and O_{R2} of P_R/P_{RM}. P_{RM} transcribes basally when $RNAP$ is bound at O_{R3}. When RNAP and CI binds to O_{R3} and O_{R2}, respectively, transcription by P_{RM} occurs at enhanced rate (activated). P_R is active only when RNAP is bound at O_{R2}/O_{R1} and O_{R3} is not occupied with CI. P_{RE} is only active when RNAP and CII binds to O_2 and O_1, respectively, hence $r4$ is the only configuration for possible CI transcription. Thus, the probability of transcribable configurations at each promoter, f_{RM}^{basal}, f_{RM}^{act}, f_R and f_{RE}, can be represented as

$$\begin{aligned} f_{RM}^{basal} &= f_{10} + f_{13} + f_{15} + f_{23} + f_{24} + f_{26} + f_{39}\,, \\ f_{RM}^{act} &= f_9 + f_{12} + f_{40}\,, \\ f_R &= f_{14} + f_{15} + f_{25}\,, \\ f_{RE} &= g_4\,. \end{aligned} \qquad (132)$$

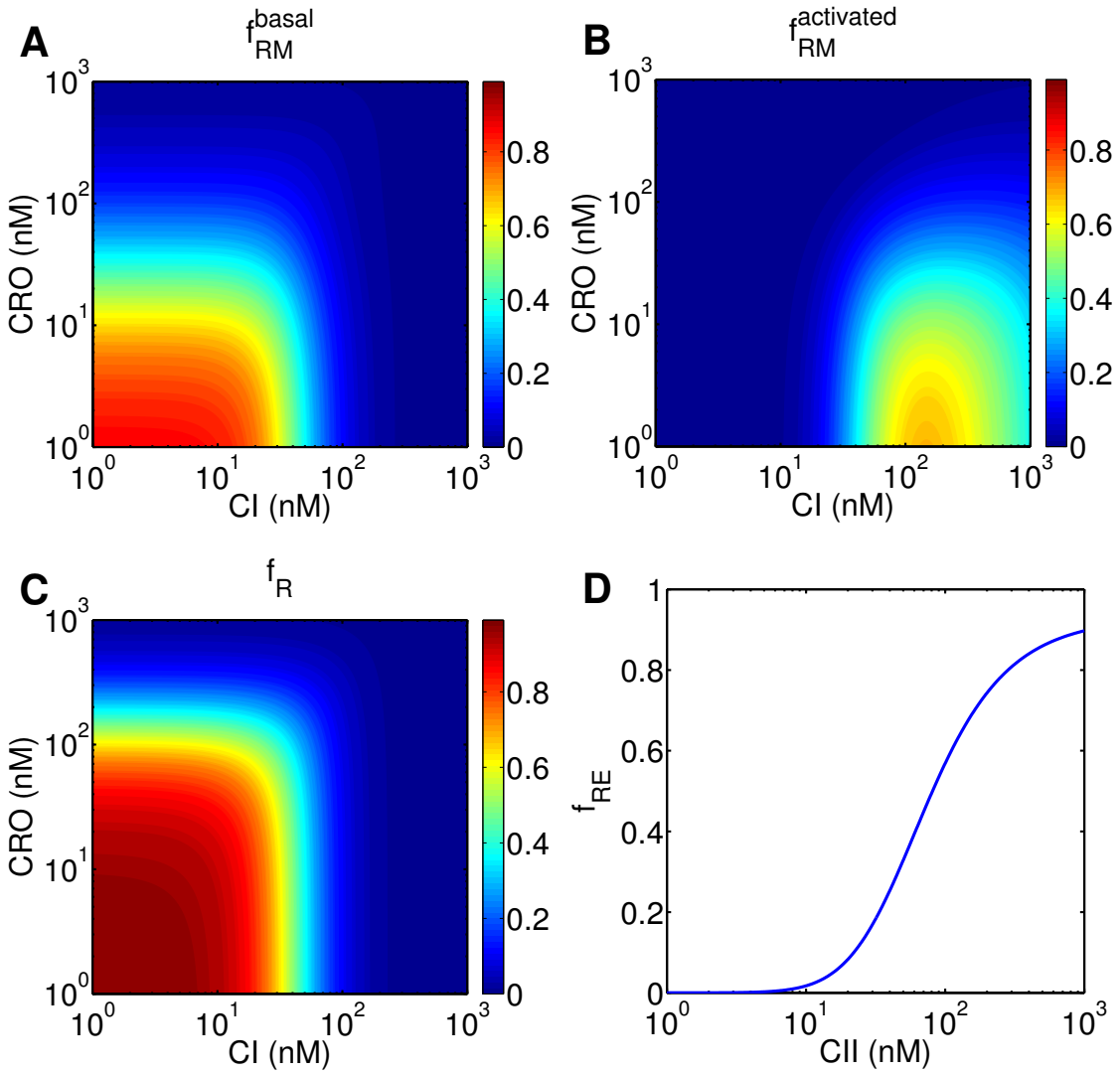

Figure 22: Probability of transcription initiation by (A) basal and (B) activated P_{RM}, (C) P_R and (D) P_{RE} as functions of total transcription factor concentrations. Note that f_{RM} and f_R is a function of CI and CRO dimer concentrations whereas f_{RE} is a function of CII tetramer concentration. By using quasi-steady-state approximation among monomers, dimers and tetramers, a fixed value of total concentration denotes a unique value of dimer/tetramer concentration.

We assume the concentration of RNAP within a cell is constant at 30 nM. The probabilities of activation for P_R, P_{RM} and P_{RE} as functions of total concentrations of CI and Q are shown in Figure 22.

Table 3: Configurations of P_R/P_{RM} and their total free energies [163].

State	O_{R3}	O_{R2}	O_{R1}	Free Energy (kcal)
$s1$	$\emptyset$	$\emptyset$	$\emptyset$	0
$s2$	$\emptyset$	$\emptyset$	CI	-11.7
$s3$	$\emptyset$	CI	$\emptyset$	-10.1
$s4$	CI	$\emptyset$	$\emptyset$	-10.1
$s5$	$\emptyset$	CI	CI	-23.7
$s6$	CI	$\emptyset$	CI	-21.8
$s7$	CI	CI	$\emptyset$	-22.2
$s8$	CI	CI	CI	-33.8
$s9$	RNAP	CI	$\emptyset$	-21.6
$s10$	RNAP	$\emptyset$	CI	-23.2
$s11$	CI	RNAP	RNAP	-22.6
$s12$	RNAP	CI	CI	-35.2
$s13$	RNAP	$\emptyset$	$\emptyset$	-11.5
$s14$	$\emptyset$	RNAP	RNAP	-12.5
$s15$	RNAP	RNAP	RNAP	-24.0
$s16$	$\emptyset$	$\emptyset$	CRO	-10.8
$s17$	$\emptyset$	CRO	$\emptyset$	-10.8
$s18$	CRO	$\emptyset$	$\emptyset$	-12.1
$s19$	$\emptyset$	CRO	CRO	-21.6
$s20$	CRO	$\emptyset$	CRO	-22.9
$s21$	CRO	CRO	$\emptyset$	-22.9
$s22$	CRO	CRO	CRO	-33.7
$s23$	RNAP	CRO	$\emptyset$	-22.3
$s24$	RNAP	$\emptyset$	CRO	-22.3
$s25$	CRO	RNAP	RNAP	-24.6
$s26$	RNAP	CRO	CRO	-33.1
$s27$	$\emptyset$	CRO	CI	-22.5
$s28$	$\emptyset$	CI	CRO	-20.9
$s29$	CI	$\emptyset$	CRO	-20.9
$s30$	CRO	$\emptyset$	CI	-23.8
$s31$	CI	CRO	$\emptyset$	-20.9
$s32$	CRO	CI	$\emptyset$	-22.2
$s33$	CRO	CI	CI	-35.8
$s34$	CI	CRO	CI	-32.6
$s35$	CI	CI	CRO	-33.0
$s36$	CI	CRO	CRO	-31.7
$s37$	CRO	CI	CRO	-33.0
$s38$	CRO	CRO	CI	-34.6
$s39$	RNAP	CRO	CI	-34.0
$s40$	RNAP	CI	CRO	-32.4

Table 4: Configurations of P_{RE} and free energies [80].

State	O_1	O_2	Free Energy (kcal)
r1	$\emptyset$	$\emptyset$	0
r2	$\emptyset$	RNAP	-9.9
r3	CII	$\emptyset$	-9.7
r4	CII	RNAP	-21.5

B.2 Stochastic simulation of phage λ switch

Eq. (3) represents the quantitative model of phage λ decision switch. Such a system of ordinary differential equations based on first order reaction kinetics can be turned into a stochastic model by using the Monte Carlo algorithm described by Gillespie [61]. All the reactions are listed in Table 5 (see Section 4.4 for definitions of parameters). Note that f_{RM}^{basal}, f_{RM}^{act} and f_R are functions of CI and CRO dimer concentrations while f_{RE} and f_{aQ} are functions of CII tetramer concentration. To demonstrate the response to viral concentration ($\mathcal{M}/V$), we vary V from 0.5 to $2\mu m^3$. V is an explicit parameter within reaction rates, and it also indirectly affects the system dynamics since the concentration change by a transcription, translation and degradation event is linearly proportional to $1/V$.

B.3 Effect of thresholds on decision making

Thresholds are key parameters within first passage process models. In this section, we show how decision thresholds affect the functional and temporal characteristics of cell fate decisions. Recall that CI and Q threshold concentrations are associated with binding affinity of transcription factors and probability of forming a DNA loop [187], and may differ between viral strains. For the first set of analysis in Results the thresholds were set at 100 nM for both CI and Q (later we modified this when considering the partial gene dosage compensation mechanism). The fraction of lysogeny as a function of $\mathcal{M}$ can be changed by tuning thresholds (see Figure 23(A) for $\mathcal{M} = 1$). We find that the fraction of lysogeny increases for all $\mathcal{M}$ by lowering the CI threshold and

Table 5: Stochastic reactions of transcription, translation and degradation. X, Y, Z and Q represent total concentrations of CI, CRO, CII and Q, respectively where as m_x, m_y, m_z, m_Q and m_{aQ} are mRNA concentrations. V is the volume of the infected host cell. See Section 4.4 for description of other parameters

Type of reactions	Interacting species and total reaction rate		
Basal cI transcription by P_{RM}	cI	$\xrightarrow{\mathcal{M}\alpha_x f_{RM}^{basal}}$	$cI + mRNA_{cI}$
Activated cI transcription by P_{RM}	cI	$\xrightarrow{\mathcal{M}\beta_x f_{RM}^{act}}$	$cI + mRNA_{cI}$
cI transcription by P_{RE}	cI	$\xrightarrow{\mathcal{M}\delta_x f_{RE}}$	$cI + mRNA_{cI}$
cro transcription	cro	$\xrightarrow{\mathcal{M}\alpha_y f_R}$	$cro + mRNA_{cro}$
cII transcription	cII	$\xrightarrow{\mathcal{M}\alpha_z f_R}$	$Q + mRNA_{cII}$
Q transcription	Q	$\xrightarrow{\mathcal{M}\alpha_Q f_R}$	$Q + mRNA_Q$
aQ transcription	aQ	$\xrightarrow{\mathcal{M}\delta_{aQ} f_{aQ}}$	$aQ + mRNA_{aQ}$
$mRNA_{cI}$ degradation	$mRNA_{cI}$	$\xrightarrow{\gamma_m m_x V}$	$\emptyset$
$mRNA_{cro}$ degradation	$mRNA_{cro}$	$\xrightarrow{\gamma_m m_y V}$	$\emptyset$
$mRNA_{cII}$ degradation	$mRNA_{cII}$	$\xrightarrow{\gamma_m m_z V}$	$\emptyset$
$mRNA_Q$ degradation	$mRNA_Q$	$\xrightarrow{\gamma_m m_Q V}$	$\emptyset$
$mRNA_{aQ}$ degradation	$mRNA_{aQ}$	$\xrightarrow{\gamma_m m_{aQ} V}$	$\emptyset$
$mRNA_Q$ degradation by $mRNA_{aQ}$	$mRNA_Q + mRNA_{aQ}$	$\xrightarrow{\zeta m_Q m_{aQ} V}$	$\emptyset$
CI degradation	CI	$\xrightarrow{\gamma_x X V}$	$\emptyset$
CRO degradation	CRO	$\xrightarrow{\gamma_y Y V}$	$\emptyset$
CII degradation	CII	$\xrightarrow{\gamma_z Z V}$	$\emptyset$
Q degradation	Q	$\xrightarrow{\gamma_Q Q V}$	$\emptyset$
CI translation	$mRNA_{cI}$	$\xrightarrow{\sigma m_x V}$	$mRNA_{cI}$+CI
CRO translation	$mRNA_{cro}$	$\xrightarrow{\sigma m_y V}$	$mRNA_{cro}$+CRO
CII translation	$mRNA_{cII}$	$\xrightarrow{\sigma m_z V}$	$mRNA_{cII}$+CII
Q translation	$mRNA_Q$	$\xrightarrow{\sigma m_Q V}$	$mRNA_Q$+Q

increasing the Q threshold for a given $\mathcal{M}/V$. On the contrary, a lower Q threshold or higher CI threshold leads to a smaller fraction of lysogeny. Note that the response to viral genome concentration is determined by the fractions of lysogeny across various $\mathcal{M}/V$, and thresholds change the fraction of lysogeny for all $\mathcal{M}/V$. Thus, even if decisions can be tuned to be almost lytic or lysogenic at a given $\mathcal{M}/V$, there might be a limit of how sensitive the fraction of lysogeny can be to the change of the viral genome concentration. So far we mainly focused on the fraction of lysogeny, the

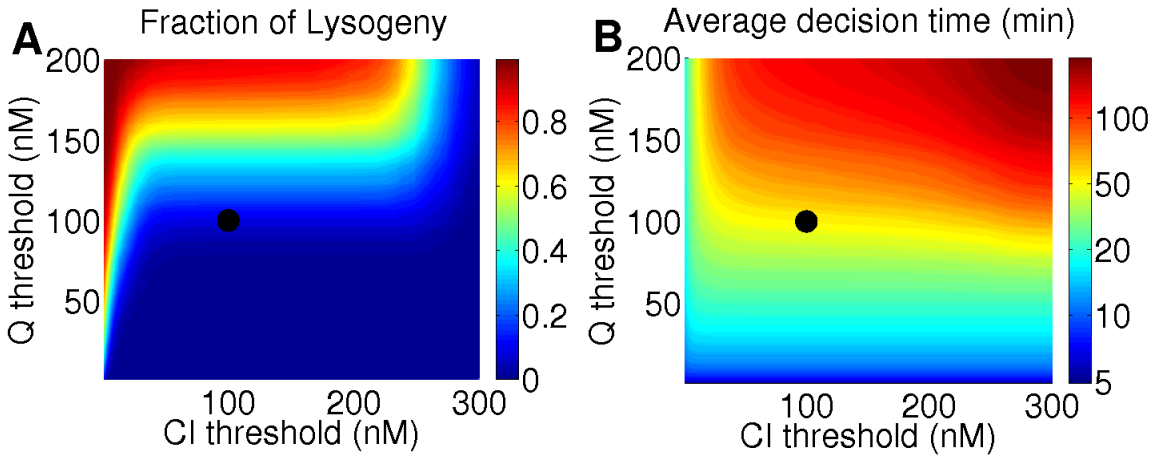

Figure 23: Effect of thresholds on decision making (A) Functional (fraction of lysogeny) and (B) temporal (mean decision time) effect of thresholds at $\mathcal{M} = 1$ when transiently divergent. Black circle shows CI and Q threshold for simulation in the previous section.

functional aspect of alternative decision making, but threshold concentrations also determine the speed at which decisions are made. Given fixed kinetic parameters, it takes less time to produce a small amount of viral proteins, so lower thresholds lead to faster decision time (see Figure 23(B)). Extreme values of thresholds lead to shorter decision times but also tend to produce non-heterogeneous decisions as a function of variation in extrinsic parameters. Hence, faster decisions may come at the expense of the ability of the viral GRN to bias cell fate determination as a function of the extrinsic parameter, $\mathcal{M}/V$.

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

C.1 Probability of each cell fate in the experimental data

Even if the focuses of cell fate prediction in phage infected bacteria are lysis and lysogeny, some cells does not develop any signs of phage infections. Here we denote such cells as naive which accounts for 19% of cells in the dataset. This can be due to host defense mechanisms [16] or defective phages. The probability of each cell fate is shown in Figure 24(A,B,C). Naive cells are associated with low levels of pRE-RFP and pR-GFP which is expected gene expression without any infecting phages. During the experiments, host cells can be infected by multiple phages, and simultaneous infection is assured by inhibiting DNA injection from phages. The number of coinfecting phages varies 1 to 5 or more per an infected cell, and fractions of multiple coinfections are shown in Table 6.

C.2 Probability of lysis and lysogeny from the mechanistic model

We applied our methods to the simulations of lysis-lysogeny decision switch [87]. The probability of lysis and lysogeny is shown in Figure 25(A,B) showing that lysis and lysogeny is associated with high expression level of Q and CI, respectively.

Table 6: Fractions of multiple coinfections in the experimental dataset. $\mathcal{M}$ denotes the number of coinfecting phages.

Number of coinfections	$\mathcal{M}=1$	$\mathcal{M}=2$	$\mathcal{M}=3$	$\mathcal{M}=4$	$\mathcal{M}=5$
Fraction	0.39	0.25	0.13	0.09	0.14

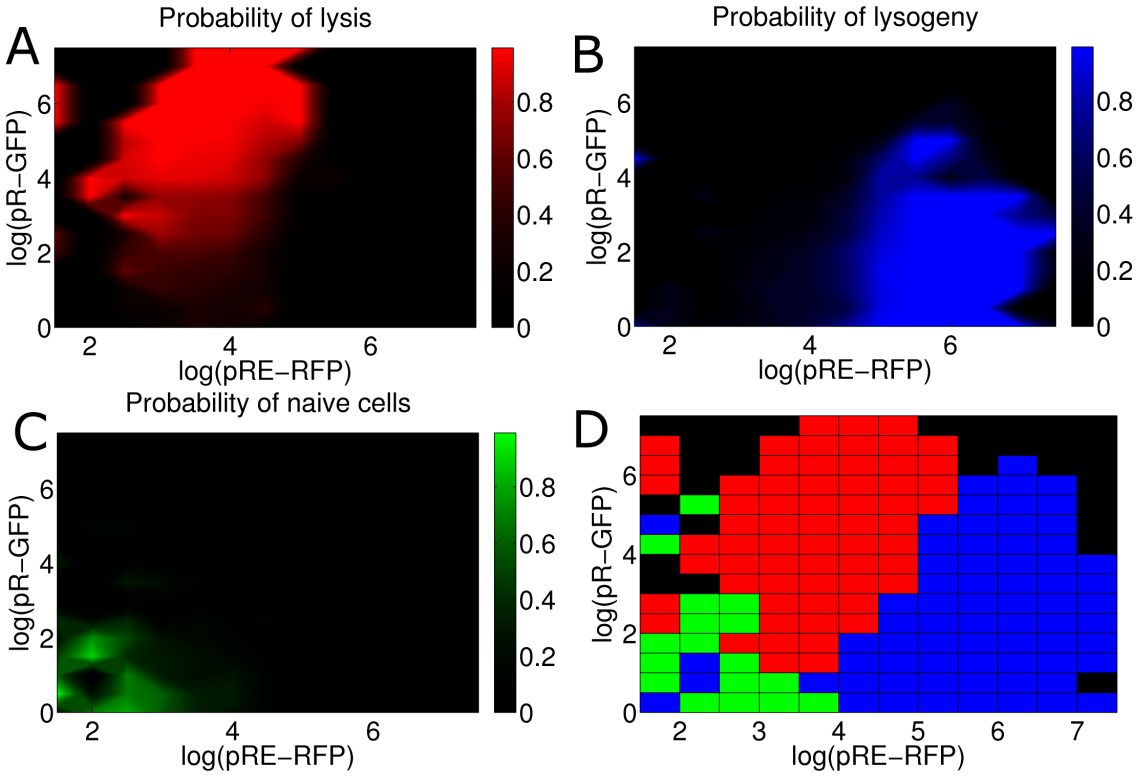

Figure 24: Probability of each cell fate from experiments. Probabilities of (A) lysis, (B) lysogeny and (C) naive cells as function of pRE-RFP and pR-GFP reporter expression level. (D) Prediction of cell fates by temporal dynamics of gene expression. The colored map represents the most likely fate in each bin of gene expression. Red, blue and green regions denote gene expression predicted with lysis, lysogeny and naive cells, respectively. Black region represents expression with no data.

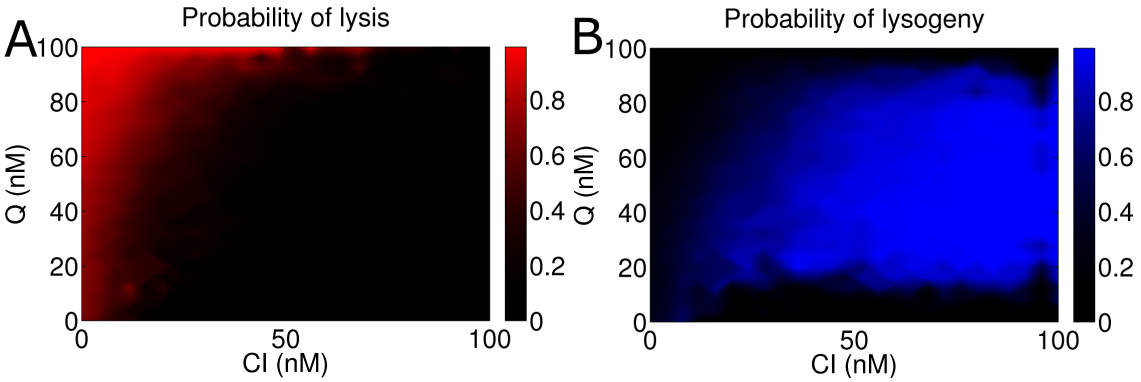

Figure 25: Probability of (A) lysis and (B) lysogeny as a function of gene expression from the quantitative model [87]. CI leads to lysogeny while Q stimulates lysis.

REFERENCES

[1] Abbott, J. D. and Shannon, R., "A method for typing shigella sonnei, using colicine production as a marker," *J. Clin. Pathol.*, vol. 11, no. 1, pp. 71–77, 1958.

[2] Ackers, G. K. and Shea, M. A., "Quantitative model for gene regulation by lambda phage repressor," *Proc. Natl. Acad. Sci. USA*, vol. 79, no. 4, pp. 1129–1133, 1982.

[3] Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[4] Alon, U., *An introduction to systems biology: Design principles of biological circuits*. Boca Raton, FL: Chapman & Hall/CRC, 2007.

[5] Amir, A., Kobiler, O., Rokney, A., Oppenheim, A., and Stavans, J., "Noise in timing and precision of gene activities in a genetic cascade," *Mol. Syst. Biol.*, vol. 3, p. 71, 2007.

[6] Anderson, R. P. and Roth, J. R., "Gene duplication in bacteria: alteration of gene dosage by sister-chromosome exchanges," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 43, pp. 1083–1087, 1979.

[7] Andersson, A., Pelve, E., Lindeberg, S., Lundgren, M., Nilsson, P., and Bernander, R., "Replication-biased genome organisation in the crenarchaeon *Sulfolobus*," *BMC Genom.*, vol. 11, no. 1, p. 454, 2010.

[8] Arkin, A., Ross, J., and McAdams, H. H., "Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells," *Genetics*, vol. 149, pp. 1633–1648, 1998.

[9] Atkinson, M. R., Savageau, M. A., Myers, J. T., and Ninfa, A. J., "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*," *Cell*, vol. 113, pp. 597–607, 2003.

[10] Atsumi, S. and Little, J., "A synthetic phage λ regulatory circuit," *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 19045–19050, 2006.

[11] Aurell, E. and Sneppen, K., "Epigenetics as a first exit problem," *Phys. Rev. Lett.*, vol. 88, p. 048101, 2002.

[12] Avlund, M., Dodd, I. B., Semsey, S., Sneppen, K., and Krishna, S., "Why do phage play dice?," *J. Virol.*, vol. 83, no. 22, pp. 11416–11420, 2009.

[13] Babić, A. C. and Little, J. W., "Cooperative DNA binding by CI repressor is dispensable in a phage λ variant," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 17741–17746, 2007.

[14] Baek, K., Svenningsen, S., Eisen, H., Sneppen, K., and Brown, S., "Single-cell analysis of λ immunity regulation," *J. Mol. Biol.*, vol. 334, pp. 363–372, 2003.

[15] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S., "Bacterial persistence as a phenotypic switch," *Science*, vol. 305, no. 5690, pp. 1622–1625, 2004.

[16] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P., "CRISPR provides acquired resistance against viruses in prokaryotes," *Science*, vol. 315, no. 5819, pp. 1709–1712, 2007.

[17] Beckmann, J. S., Estivill, X., and Antonarakis, S. E., "Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability," *Nat. Rev. Genet.*, vol. 8, pp. 639–646, 2007.

[18] Becskei, A., Seraphin, B., and Serrano, L., "Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion," *EMBO J.*, vol. 20, pp. 2528–2535, 2001.

[19] Ben-Israel, H. and Kleinberger, T., "Adenovirus and cell cycle control," *Front. Biosci.*, vol. 7, pp. d1369–d1395, 2002.

[20] Bennett, M., Volfson, D., Tsimring, L., and Hasty, J., "Transient dynamics of genetic regulatory networks," *Biophys. J.*, vol. 92, pp. 3501–3512, 2007.

[21] Birchler, J. A., Riddle, N. C., Auger, D. L., and Veitia, R. A., "Dosage balance in gene regulation: biological implications," *Trends Genet.*, vol. 21, pp. 219–226, 2005.

[22] Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, K., Wilson, M. J., Rossant, J., Hamada, H., and Koopman, P., "Retinoid signaling determines germ cell fate in mice," *Science*, vol. 312, no. 5773, pp. 596–600, 2006.

[23] Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., and Helin, K., "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions," *Genes Dev.*, vol. 20, pp. 1123–1136, 2006.

[24] Brookfield, J. F. Y. and Badge, R. M., "Population genetics models of transposable elements," *Genetica*, vol. 100, pp. 281–294, 1997.

[25] Burz, D., Becket, D., Benson, N., and Ackers, G. K., "Self-assembly of bacteriophage λ CI repressor: effects of single-site mutations on the monomer-dimer equilibrium," *Biochemistry*, vol. 33, pp. 8399–8405, 1994.

[26] Calendar, R., ed., *The Bacteriophages*. Oxford University Press USA, 2005.

[27] Camazine, S., Sneyd, J., and Seeley, T., "Collective decision-making in honey-bees - how colonies choose among nectar sources," *Behav. Ecol. Sociobiol.*, vol. 28, pp. 277–290, 1991.

[28] Cantrell, D. A. and Smith, K. A., "Transient expression of interleukin 2 receptors. Consequences for T cell growth," *J. Exp. Med.*, vol. 158, no. 6, pp. 1895–1911, 1983.

[29] Casjens, S., "Prophages and bacterial genomics: what have we learned so far?," *Mol. Microbiol.*, vol. 49, no. 2, pp. 277–300, 2003.

[30] Cayouette, M., Barres, B., and Raff, M., "Importance of intrinsic mechanisms in cell fate decisions in the developing rat retina," *Neuron*, vol. 40, p. 897904, 2003.

[31] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C., "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, 1994.

[32] Chan, J., Stampfer, M., Giovannucci, E., Gann, P., Ma, J., Wilkinson, P., Hennekens, C., and Pollak, M., "Plasma insulin-like growth factor i and prostate cancer risk: A prospective study," *Science*, vol. 279, no. 5350, pp. 563–566, 1998.

[33] Cherry, J. L. and Adler, F. R., "How to make a biological switch," *J. Theor. Biol.*, vol. 203, no. 2, pp. 117 – 133, 2000.

[34] Chittka, L., Dyer, A. G., Bock, F., and Dornhaus, A., "Psychophysics: bees trade off foraging speed for accuracy," *Nature*, vol. 424, p. 388, 2003.

[35] Ciliberti, S., Martin, O. C., and Wagner, A., "Robustness can evolve gradually in complex regulatory gene networks with varying topology," *PLoS Comp. Biol.*, vol. 3, p. e15, 2007.

[36] Cohen, A. R., Gomes, F. L. A. F., Roysam, B., and Cayouette, M., "Computational prediction of neural progenitor cell fates," *Nat. Meth.*, vol. 7, pp. 213–218, 2010.

[37] Conrad, B. and Antonarakis, S. E., "Gene duplication: a drive for phenotypic diversity and cause of human disease," *Annu. Rev. Genomics Hum. Genet.*, vol. 8, pp. 17–35, 2007.

[38] Cooper, G. M., Nickerson, D. A., and Eichler, E. E., "Mutational and selective effects on copy-number variants in the human genome," *Nat. Genet.*, vol. 39, pp. S22–S29, 2007.

[39] Court, D., Crombrugghe, B., Adhya, S., and Gottesman, M., "Bacteriophage lambda Hin function II. Enhanced stability of lambda messenger RNA," *J. Mol. Biol.*, vol. 138, pp. 731–743, 1980.

[40] Court, D. L., Oppenheim, A. B., and Adhya, S. L., "A new look at bacteriophage λ genetic networks," *J. Bacteriol.*, vol. 189, no. 2, pp. 298–304, 2007.

[41] Couzin, I., Krause, J., Franks, N., and Levin, S., "Effective leadership and decision-making in animal groups on the move," *Nature*, vol. 433, pp. 513–516, 2005.

[42] Darling, D. A. and Siegert, A. J. F., "The first passage problem for a continuous Markov process," *Ann. Math. Statist.*, vol. 4, pp. 624–639, 1953.

[43] Darling, P. J., Holt, J. M., and Ackers, G. K., "Couple energetics of λ *cro* repressor self-assembly and site-specific DNA operator binding. I: Analysis of *cro* dimerization from nanomolar to micromolar concentrations," *Biochemistry*, vol. 39, pp. 11500–11507, 2000.

[44] Davidson, E. H. and Erwin, D. H., "Gene regulatory networks and the evolution of animal body plans," *Science*, vol. 311, pp. 796–800, 2006.

[45] De Paepe, M. and Taddei, F., "Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages," *PLoS Biol.*, vol. 4, p. e193, 2006.

[46] Dean, E. J., Davis, J. C., Davis, R. W., and Petrov, D. A., "Pervasive and persistent redundancy among duplicated genes in yeast," *PLoS Genet.*, vol. 4, p. e1000113, 2008.

[47] Deluna, A., Vetsigian, K., Shoresh, N., Hegreness, M., Colón-González, M., Chao, S., and Kishony, R., "Exposing the fitness contribution of duplicated genes," *Nat. Genet.*, vol. 40, pp. 676–681, 2008.

[48] Dimmock, N., Easton, A., and Leppard, K., *Introduction to Modern Virology*. Malden, MA: Blackwell Publishing, 5th ed., 2007.

[49] Dodd, I. B., Shearwin, K. E., and Egan, J. B., "Revisited gene regulation in bacteriophage λ," *Curr. Opin. Genes Dev.*, vol. 15, no. 2, pp. 145 – 152, 2005.

[50] Ebisuya, M., Kondoh, K., and Nishida, E., "The duration, magnitude and compartmentalization of EKR MAP kinase activity: mechanisms for providing signaling specificity," *J. Cell Sci.*, vol. 118, pp. 2997–3002, 2005.

[51] Edgar, R., Rokney, A., Feeney, M., Semsey, S., Kessel, M., Goldberg, M. B., Adhya, S., and Oppenheim, A. B., "Bacteriophage infection is targeted to cellular poles," *Mol. Microbiol.*, vol. 68, no. 5, pp. 1107–1116, 2008.

[52] Elowitz, M. and Leibler, S., "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, pp. 335–338, 2000.

[53] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S., "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, 2002.

[54] Endy, D., "Foundations for engineering biology," *Nature*, vol. 438, pp. 449–453, 2005.

[55] Eraso, P., Cid, A., and Serrano, R., "Tight control of the amount of yeast plasma membrane atpase during changes in growth conditions and gene dosage," *FEBS Lett.*, vol. 224, no. 1, pp. 193 – 197, 1987.

[56] Fabris, C., Valduga, G., Miotto, G., Borsetto, L., Jori, G., Garbisa, S., and Reddi, E., "Photosensitization with Zinc (II) phthalocyanine as a switch in the decision between apoptosis and necrosis," *Cancer Res.*, vol. 61, no. 20, pp. 7495–7500, 2001.

[57] Gardner, T. S., Cantor, C. R., and Collins, J. J., "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, pp. 339–342, 2000.

[58] Germain, R. N., "T-cell development and the CD4-CD8 lineage decision," *Nat. Rev. Immunol.*, vol. 2, pp. 209–322, 2002.

[59] Gerweck, L. E. and Seetharaman, K., "Cellular pH gradient in tumor versus normal tissue: potential exploitation for the treatment of cancer," *Cancer Res.*, vol. 56, no. 6, pp. 1194–1198, 1996.

[60] Gett, A. V., Sallusto, F., Lanzavechhia, A., and Genginat, J., "T cell fitness determined by signal strength," *Nat. Immunol.*, vol. 4, pp. 355–360, 2003.

[61] Gillespie, D. T., "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 82, no. 25, pp. 2340–2361, 1977.

[62] Greenblatt, J., Nodwell, J. R., and Mason, S. W., "Transcriptional antitermination," *Nature*, vol. 364, no. 6436, pp. 401–406, 1993.

[63] Griffith, J., Hochschild, A., and Ptashne, M., "DNA loops induced by cooperative binding of λ repressor," *Nature*, vol. 322, pp. 750–752, 1986.

[64] Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., and Carroll, S. B., *Introduction to Genetic Analysis*. New York, NY: W. H. Freeman and Company, 2008.

[65] Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J.-E., Kawakami, K., Xu, P.-X., Kelly, R., Petrof, B. J., Daegelen, D., Concordet, J.-P., and Maire, P., "Six1 and eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype," *Mol. Cell. Biol.*, vol. 24, no. 14, pp. 6253–6267, 2004.

[66] Guan, Y., Dunham, M. J., and Troyanskaya, O. G., "Functional analysis of gene duplications in *Saccharomyces cerevisiae*," *Genetics*, vol. 175, pp. 933–943, 2007.

[67] Gudelj, I., Weitz, J. S., Ferenci, T., Horner-Devine, M. C., Marx, C. J., Meyer, J. R., and Forde, S. E., "An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure," *Ecol. Lett.*, vol. 13, pp. 1073–1084, 2010.

[68] Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., Elston, T. C., and Collins, J. J., "A bottom-up approach to gene regulation," *Nature*, vol. 439, pp. 856–860, 2006).

[69] Håkelien, A.-M., Gaustad, K. G., and Collas, P., "Transient alteration of cell fate using a nuclear and cytoplasmic extract of an insulinoma cell line," *Biochem. Biophys. Res. Commun.*, vol. 316, no. 3, pp. 834 – 841, 2004.

[70] Hagiwara, S., Miyazaki, S., and Rosenthal, N. P., "Potassium current and the effect of cesium on this current during anomalous rectification of the egg cell membrane of a starfish," *J. Gen. Physiol.*, vol. 67, pp. 1621–658, 1976.

[71] Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J., "Computational studies of gene regulatory networks: *in numero* molecular biology," *Nat. Rev. Genet.*, vol. 2, pp. 268–279, 2001.

[72] Hawley, D. K., Johnson, A. D., and McClure, W. R., "Functional and physical characterization of transcription initiation complexes in the bacteriophage λ O_R region," *J. Biol. Chem.*, vol. 260, pp. 8618–8626, 1985.

[73] Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R. A., *Lambda II*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1983.

[74] Herskowitz, I. and Hagen, D., "The lysis-lysogeny decision of phage λ: explicit programming and responsiveness," *Ann. Rev. Genet.*, vol. 14, pp. 399–445, 1980.

[75] Hitoshi, S., Tropepe, V., Ekker, M., and van der Kooy, D., "Neural stem cell lineages are regionally specified, but not committed, within distinct compartments of the developing brain," *Development*, vol. 129, no. 1, pp. 233–244, 2002.

[76] Hittinger, C. T. and Carroll, S. B., "Gene duplication and the adaptive evolution of a classic genetic switch," *Nature*, vol. 449, pp. 677–681, 2007.

[77] Hochschild, A. and Ptashne, M., "Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix," *Cell*, vol. 44, pp. 681–687, 1984.

[78] Hockett, R. D., Kilby, M. J., Derdeyn, C. A., Saag, M. S., Sillers, M., Suires, K., Chiz, S., Norwak, M. A., Shaw, G. M., and Bucy, P. R., "Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA," *J. Exp. Med.*, vol. 189, no. 10, 1999.

[79] Howes, W. V., "Protein synthesis in *Escherichia coli*: a phage mediated interruption of tranalasion," *Biochim. Biophys. Acta*, vol. 103, pp. 711–713, 1965.

[80] Hoyt, M. A., Knight, D. M., Das, A., Miller, H. I., and Echols, H., "Control of phage lambda development by stability and synthesis of *cii* protein: role of the viral *cii* and host $hfla$, *hima* and *himd* genes," *Cell*, vol. 31, pp. 565–573, 1982.

[81] Hwang, J. J., Brosn, S., and Gussin, G. N., "Characterization of a doubly mutant derivative of the lambda P_{RM} promoter. Effects of mutations on activation of P_{RM}," *J. Mol. Biol.*, vol. 200, pp. 695–708, 1988.

[82] Ingolia, N. T. and Murray, A. W., "Positive-feedback loops as a flexible biological module," *Curr. Biol.*, vol. 17, pp. 668–677, 2007.

[83] Isaacs, F. J., Hasty, J., Cantor, C. R., and Collins, J. J., "Prediction and measurement of an autoregulatory genetic module," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 7714–7719, 2003.

[84] Jakobsson, M., Scholz, S. W., Scheet, P., Gibbs, J. R., VanLiere, Jenna M.and Fung, H.-C., Szpiech, Z. A., Degnan, J. H., Wang, K., Guerreiro, R., Bras, J. M., Schymick, J. C., Hernandez, D. G., Traynor, B. J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H. M., Hardy, J. A., Rosenberg, N. A., and Singleton, A. B., "Genotype, haplotype and copy-number variation in worldwide human populations," *Nature*, vol. 451, no. 7181, pp. 998–1003, 2008.

[85] Jana, R., Hazbun, T. R., Mollah, A. K., and Mossing, M. C., "A folded monomeric intermediate in the formation of lambda Cro dimer-dna complexes," *J. Mol. Biol.*, vol. 273, pp. 402–416, 1997.

[86] Joh, R. I., Lee, T., and Weitz, J. S., "The future is now: predicting cell fate by measuring temporal dynamics of gene regulation." in prep., 2011.

[87] Joh, R. I. and Weitz, J. S., "To lyse or not to lyse: transient-mediated stochastic fate determination in cells infected by bacteriophages," *PLos Comp. Biol.*, vol. 7, p. e1002006, 2011.

[88] Johnson, G. L. and Vaillancourt, R. R., "Sequential protein-kinase reactions controlling cell-growth and differentiation," *Curr. Opin. Cell Biol.*, vol. 6, pp. 230–238, APR 1994.

[89] Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J., "Stochasticity in gene expression: from theories to phenotypes," *Nat. Rev. Genet.*, vol. 6, pp. 451–464, 2005.

[90] Kaufmann, B. B. and van Oudenaarden, A., "Stochastic gene expression: from single molecules to the proteome," *Curr. Opin. Genet. Dev.*, vol. 17, pp. 107–112, 2007.

[91] Keller, G. M., "In vitro differentiation of embryonic stem cells," *Curr. Opin. Cell Biol.*, vol. 7, no. 6, pp. 862 – 869, 1995.

[92] Kellis, M., Birren, B. W., and Lander, E. S., "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 428, pp. 617–624, 2004.

[93] Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N. A., Tsang, P., Newman, T. L., Tuzun, E., Cheng, Z., Ebling, H. M., Tusneem, N., David, R., Gillett, W., Phelps, K. A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J. D., Korn, J. M., McCarroll, S. A., Altshuler, D. A., Peiffer, D. A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D. A., Mullikin, J. C., Wilson, R. K., Bruhn, L., Olson, M. V., Kaul, R., Smith, D. R., and Eichler, E. E., "Mapping and sequencing of structural variation from eight human genomes," *Nature*, vol. 453, pp. 56–64, 2008.

[94] Kihara, A., Akiyama, Y., and Ito, K., "Host regulation of lysogenic decision in bacteriophage λ: Transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC(HflA)," *Proc. Natl. Acad. Sci. USA*, vol. 94, no. 11, pp. 5544–5549, 1997.

[95] Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M., "rRNA oepron copy number reflects ecological strategies of bacteria," *Appl. Environ. Microbiol.*, vol. 66, pp. 1328–1333, 2000.

[96] Kobiler, O., Koby, S., Teff, D., Court, D., and Oppenheim, A. B., "The phage λ CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 23, pp. 14964–14969, 2002.

[97] Kobiler, O., Rokney, A., Fredman, N., Court, D. L., Stravans, J., and Oppenheim, A. B., "Quantitative kinetic analysis of the bacteriophage λ genetic network," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 12, pp. 4470–4475, 2005.

[98] Kobiler, O., Rokney, A., and Oppenheim, A., "Phage lambda CIII: a protease inhibitor regulating the lysis-lysogeny decision," *PLoS One*, vol. 2, p. e363, 2007.

[99] Kollmann, M., Lovdok, L., Bartholome, K., Timmer, J., and Sourjik, V., "Design principles of a bacterial signalling network," *Nature*, vol. 438, pp. 504–507, 2005.

[100] Korbel, J. O., Kim, P. M., Chen, X., Urban, A. E., Weissman, S., Snyder, M., and Gerstein, M. B., "The current excitement about copy-number variation: how it relates to gene duplications and protein families," *Curr. Opin. Struct. Biol.*, vol. 18, pp. 366–374, 2008.

[101] Kourilsky, P., "Lysogenization by bacteriophage lambda. I. Multiple infection and the lysogenic response," *Mol. Gen. Genet.*, vol. 122, pp. 183–195, 1973.

[102] Kussell, E. and Leibler, S., "Phenotypic diversity, population growth, and information in fluctuating environments," *Science*, vol. 309, pp. 2075–2078, 2005.

[103] Laurent, M. and Kellershohn, N., "Multistability: a major means of differentiation and evolution in biological systems," *Trends Biochem. Sci.*, vol. 24, no. 11, pp. 418–422, 1999.

[104] Lee, C.-K., Klopp, R. G., Weindruch, R., and Prolla, T. A., "Gene expression profile of aging and its retardation by caloric restriction," *Science*, vol. 285, no. 5432, pp. 1390–1393, 1999.

[105] Lee, T. J., Yao, G., Bennett, D. C., Nevins, J. R., and You, L., "Stochastic E2F activation and reconciliation of phenomenological cell-cycle models," *PLoS Biol*, vol. 8, p. e1000488, 09 2010.

[106] Levine, E. and Hwa, T., "Small RNAs establish gene expression thresholds," *Curr. Opin. Microbiol.*, vol. 11, pp. 574–579, 2008.

[107] Li, W. H., Yang, J., and Gu, X., "Expression divergence between duplicate genes," *Trends Gen.*, vol. 21, pp. 602–607, 2005.

[108] Lipshtat, A., Loinger, A., Balaban, N. Q., and Biham, O., "Genetic toggle switch without cooperative binding," *Phys. Rev. Lett.*, vol. 96, no. 18, p. 188101, 2006.

[109] Little, J., Shepley, D., and Wert, D., "Robustness of a gene regulatory circuit," *EMBO J.*, vol. 18, pp. 4299–4307, 1999.

[110] Losick, R. and Desplan, C., "Stochasticity and cell fate," *Science*, vol. 320, no. 5872, pp. 65–68, 2008.

[111] Lynch, M. and Conery, J. S., "The evolutionary fate and consequences of duplicate genes," *Science*, vol. 290, pp. 1151–1155, 2000.

[112] Maamar, H. and Dubnau, D., "Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop," *Mol. Microbiol.*, vol. 56, no. 3, pp. 615–624, 2005.

[113] Maamar, H., Raj, A., and Dubnau, D., "Noise in gene expression determines cell in *Bacillus subtilis*," *Science*, vol. 317, pp. 526–529, 2007.

[114] Maldonado, R., Jimenez, J., and Casadesus, J., "Changes of ploidy during the azotobacter vinelandii growth cycle.," *J. Bacteriol.*, vol. 176, no. 13, pp. 3911–3919, 1994.

[115] Mangan, S., Itzkovitz, S., Zaslaver, A., and Alon, U., "The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*," *J. Mol. Biol.*, vol. 356, no. 5, pp. 1073 – 1081, 2006.

[116] Mangan, S., Zaslaver, A., and Alon, U., "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks," *J. Mol. Biol.*, vol. 334, no. 2, pp. 197 – 204, 2003.

[117] Marshall, C. J., "Specificity of receptop tyrosine kinase signaling: transient versus susained extracellular signal-regulated kinase activation," *Cell*, vol. 80, no. 2, pp. 179–185, 1995.

[118] McAdams, H. H. and Arkin, A., "It's a noisy business! Genetic regulation at the nanomolar scale," *Trends Genet.*, vol. 15, pp. 65–69, 1999.

[119] McAdams, H. and Arkin, A., "Stochastic mechanisms in gene expression," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 814–819, 1997.

[120] McPherson, K., Steel, C. M., and Dixon, J. M., "Abc of breast disease: Breast cancer-epidemiology, risk factors, and genetics," *Brit. Med. J.*, vol. 321, pp. 624–628, 2000.

[121] Mehta, P., Mukhopadhyay, R., and Wingreen, N. S., "Exponential sensitivity of noise-driven switching in genetic networks," *Phys. Biol.*, vol. 5, no. 2, p. 026005, 2008.

[122] Michael, T. P., Salom, P. A., Yu, H. J., Spencer, T. R., Sharp, E. L., McPeek, M. A., Alonso, J. M., Ecker, J. R., and McClung, C. R., "Enhanced fitness conferred by naturally occurring variation in the circadian clock," *Science*, vol. 302, no. 5647, pp. 1049–1053, 2003.

[123] Mileyko, Y. and Weitz, J. S., "Bifurcation analysis of gene regulatory circuits subject to copy number variation," *SIAM J. Appl. Dyn. Syst.*, vol. 9, pp. 799–826, 2010.

[124] Mileyko, Y., Joh, R. I., and Weitz, J. S., "Small-scale copy number variation and large-scale changes in gene expression," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 43, pp. 16659–16664, 2008.

[125] Miller, M. and Bassler, B., "Quorum sensing in bacteria," *Annu. Rev. Microbiol.*, vol. 55, pp. 165–199, 2001.

[126] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U., "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, pp. 824–827, 2002.

[127] Mishan, E. J. and Quah, E., *Cost benefit analysis*. New York, NY: Routledge, 5th ed., 2007.

[128] Morelli, M. J., ten Wolde, P. R., and Allen, R. J., "DNA looping provides stability and robustness to the bacteriophage λ switch," *Proc. Natl. Acad. Sci. USA*, vol. 106, pp. 8101–8106, 2009.

[129] Morrison, S. J. and Shar, N. M., "Regulatory mechanisms in stem cell biology," *Cell*, vol. 88, pp. 287–298, 1997.

[130] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B., "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nat. Meth.*, vol. 5, no. 7, pp. 621–628, 2008.

[131] Murray, N. and Gann, A., "What has phage lambda ever done for us?," *Curr. Biol.*, vol. 17, pp. R305–R312, 2007.

[132] Musso, G., Costanzo, M., Huangfu, M. Q., Smith, A. M., Paw, J., Luis, B. J. S., Boone, C., Giaever, G., Nislow, C., Emili, A., and Zhang, Z. L., "The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast," *Genome Res.*, vol. 18, pp. 1092–1099, 2008.

[133] Nagpal, P., Jafri, S., Reddy, M. A., and Das, H. K., "Multiple chromosomes of *Azotobacter vinelandii*," *J. Bacteriol.*, vol. 171, no. 6, pp. 3133–3138, 1989.

[134] Nicotera, P., Leist, M., and Ferrando-May, E., "Intracellular ATP, a switch in the decision between apoptosis and necrosis," *Toxicol. Lett.*, vol. 102-103, pp. 139 – 142, 1998.

[135] Novick, S. L. and Baldeschwieler, J. D., "Fluorescence measurement of kinetics of DNA injection by bacteriophage λ into liposomes," *Biochemistry*, vol. 27, pp. 7919–7924, 1988.

[136] Ohno, S., *Evolution by Gene Duplication*. London, UK: Allen and Unwin, 1970.

[137] Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S., "Switches in bacteriophage lambda development," *Annu. Rev. Genet.*, vol. 39, no. 1, pp. 409–429, 2005.

[138] Pakula, A. A., Young, V. B., and Sauer, R. T., "Bacteriophage lambda *cro* mutations: effects on activity and intracellular degradation," *Proc. Natl. Acad. Sci. USA*, vol. 83, pp. 8829–8833, 1986.

[139] Parua, P. K., Datta, A. B., and Parrack, P., "Specific hydrophobic residues in the α4 helix of λCII are crucial for maintaining its tetrameric structure and directing the lysogenic choice," *J. Gen. Virol*, vol. 91, no. 1, pp. 306–312, 2010.

[140] Paulsson, J. and Ehrenberg, M., "Noise in a minimal regulatory network: plasmid copy number control," *Quart. Rev. Biophys.*, vol. 34, pp. 1–59, 2001.

[141] Pedraza, J. M. and van Oudenaarden, A., "Noise propagation in gene networks," *Science*, vol. 307, pp. 1965–1969, 2005.

[142] Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., and Stone, A. C., "Diet and the evolution of human amylase gene copy number variation," *Nat. Genet.*, vol. 39, p. 12561260, 2007.

[143] Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O., "Genome-wide analysis of DNA copy-number changes using cDNA microarrays," *Nat. Genet.*, vol. 23, pp. 41–46, 1999.

[144] Pollack, J. R., Sørlie, T., Perou, C. M., Rees, C. A., Jeffrey, S. S., Lonning, P. E., Tibshirani, R., Botstein, D., Børresen-Dale, A.-L., and Brown, P. O., "Microarray analysis reveals a major direct role of

DNA copy number alteration in the transcriptional program of human breast tumors," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 12963–12968, 2002.

[145] Presser, A., Elowitz, M. B., Kellis, M., and Kishony, R., "The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication," *Proc. Natl. Acad. Sci. USA*, vol. 105, pp. 950–954, 2008.

[146] Ptashne, M., *A genetic switch. Phage lambda revisted*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 3rd ed., 2004.

[147] Rani, S. A., Pitts, B., and Stewart, P. S., "Rapid diffusion of fluorescent tracers into staphylococcus epidermidis biofilms visualized by time lapse microscopy," *Antimicrob. Agents Chemother.*, vol. 49, no. 2, pp. 728–732, 2005.

[148] Redner, S., *A guide to first-passage processes*. Cambridge, UK: Cambridge University Press, 1st ed., 2001.

[149] Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W., and Hurles, M. E., "Global variation in copy number variation in the human genome," *Nature*, vol. 444, pp. 444–454, 2006.

[150] Révet, B., von Wilcken-Bergmann, B., Bessert, H., Barker, A., and Muller-Hill, B., "Four dimers of λ repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances," *Curr. Biol.*, vol. 9, no. 3, pp. 151–154, 1999.

[151] Rhind, N. R., Miller, L. M., Kopczynski, J. B., and Meyer, B. J., "*xo1-1* acts as an early switch in the *C. elegans* male/hermaphrodite decision," *Cell*, vol. 80, no. 1, pp. 71–82, 1995.

[152] Ridker, P. M., Brown, N. J., Vaughan, D. E., Harrison, D. G., and Mehta, J. L., "Established and emerging plasma biomarkers in the prediction of first atherothrombotic events," *Circulation*, vol. 106, pp. IV–6–IV–19, 2004.

[153] Rokney, A., Kobiler, O., Amir, A., Court, D., Stavans, J., Adhya, S., and Oppenheim, A., "Host responses influence on the induction of lambda prophage," *Mol. Microbiol.*, vol. 68, pp. 29–36, 2008.

[154] Roper, R. J. and Reeves, R. H., "Understanding the basis for Down Syndrome phenotypes," *PLoS Genet.*, vol. 2, p. e50, 2006.

[155] Rotem, E., Rokney, A., Feeney, M., Semsey, S., Kessel, M., Goldberg, M. B., Adhya, S., and Oppenheim, A. B., "Bacteriophage infection is targeted to cellular poles," *Mol. Microbiol.*, vol. 5, pp. 1107–1116, 2008.

[156] Russell, J. H. and Keiler, K. C., "Subcellular localization of a bacterial regulatory RNA," *Proc. Natl. Acad. Sci. USA*, vol. 106, pp. 16405–16409, 2009.

[157] Saitou, M., Barton, S. C., and Azim Surani, M., "A molecular programme for the specification of germ cell fate in mice," *Nature*, vol. 418, pp. 293–300, 2002.

[158] Sanga, S., Frieboes, H. B., Zheng, X., Gatenby, R., Bearer, E. L., and Cristini, V., "Predictive oncology: A review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth," *Neuroimage*, vol. 37, no. S1, pp. S120 – S134, 2007.

[159] Santillán, M. and Mackey, M., "Why the lysogenic state of phage λ is so stable: a mathematical modeling approach," *Biophys. J.*, vol. 86, pp. 75–84, 2004.

[160] Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M., and Backman, K. C., "Regulatory functions of the lambda. Repressor reside in the amino-terminal domain," *Nature*, vol. 279, no. 5712, pp. 396–400, 1979.

[161] Schuster-Böckler, B., Conrad, D., and Bateman, A., "Dosage sensitivity shapes the evolution of copy-number varied regions," *PLoS ONE*, vol. 5, p. e9474, 03 2010.

[162] Schweitzer, R., Chuyung, J. H., Murtaugh, L. C., Brent, A. E., Rossen, V., Olson, E. N., Lassar, A., and Tabin, C. J., "Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments," *Development*, vol. 128, pp. 3855–3866, 2001.

[163] Shea, M. A. and Ackers, G. K., "The O_R control system of bacteriophage lambda. A physical-chemical model for gene regulation," *J. Mol. Biol.*, vol. 181, no. 2, pp. 211–230, 1985.

[164] Sherratt, D. J., "Bacterial chromosome dynamics," *Science*, vol. 301, pp. 780–785, 2003.

[165] Singh, A. and Weinberger, L. S., "Stochastic gene expression as a molecular switch for viral latency," *Curr. Opin. Microbiol.*, vol. 12, no. 4, pp. 460 – 466, 2009.

[166] Slutsky, M. Mirny, L. A., "Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential," *Biophys. J.*, vol. 87, pp. 4021–4035, 2004.

[167] Smith, A. G., Heath, J. K., Donaldson, D. D., Wond, G. G., Morreau, J., Stahl, M., and Rogers, D., "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides," *Nature*, vol. 336, pp. 688–690, 1988.

[168] Solomon, J. M. and Grossman, A. D., "Who's competent and when: regulation of natural genetic competence in bacteria," *Trends Genet.*, vol. 12, no. 4, pp. 150 – 155, 1996.

[169] Springer, M., Weissman, J. S., and Kirschner, M. W., "A general lack of compensation for gene dosage in yeast," *Mol. Syst. Biol.*, vol. 6, p. 368, MAY 2010.

[170] Spudich, J. L. and Koshland Jr., D. E., "Non-genetic individuality: chance in the single cell," *Nature*, vol. 262, pp. 467–471, 1976.

[171] St. Pierre, F. and Endy, D., "Determination of cell fate selection during phage lambda infection," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 52, pp. 20705–20710, 2008.

[172] Stevenson, B. S. and Schmidt, T. M., "Life history implications of rRNA gene copy number in *Escherichia coli*," *Appl. Environ. Microbiol.*, vol. 70, pp. 6670–6677, 2004.

[173] Stewart, F. M. and Levin, B. R., "The population biology of bacterial viruses: Why be temperate?," *Theor. Pop. Biol.*, vol. 26, pp. 93–117, 1984.

[174] Stragier, P. and Losick, R., "Molecular genetics of sporulation in *Bacillus subtilis*," *Annu. Rev. Gen.*, vol. 30, no. 1, pp. 297–341, 1996.

[175] Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavar, S., Deloukas, P., Hurles, M. E., and Dermitzakis, E. T., "Relative impact of nucleotide and copy number variation on gene expression phenotypes," *Science*, vol. 315, no. 5813, pp. 848–853, 2007.

[176] Strogatz, S., *Nonlinear Dynamics and Chaos*. Reading, MA: Addison Wesley, 1994.

[177] Süel, M. B., Garcia-Ojalvo, J., Liberman, L. M., and Elowitz, M. B., "An excitable gene regulatory circuit induces transient cellular differentiation," *Nature*, vol. 440, pp. 545–550, 2006.

[178] Svenningsen, S. L., Tu, K. C., and Bassler, B. L., "Gene dosage compensation calibrates four regulatory RNAs to control *Vibriio cholerae* quorum sensing," *EMBO J.*, vol. 28, pp. 429–439, 2009.

[179] Swat, M., Kel, A., and Herzel, H., "Bifurcation analysis of the regulatory modules of the mammalian G(1)/S transition," *Bioinformatics*, vol. 20, no. 10, pp. 1506–1511, 2004.

[180] Takizawa, P. A., Sil, A., Swedlow, J. R., Herskowitz, I., and Vale, R. D., "Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast," *Nature*, vol. 389, pp. 90–93, 1987.

[181] Terzi, M. and Levinthal, C., "Effects of λ-phage infection on bacterial synthesis," *J. Mol. Biol.*, vol. 26, no. 3, pp. 525 – 530, 1967.

[182] Thanbichler, M. and Shapiro, L., "Getting organized-how bacterial cells move proteins and DNA," *Nat. Rev. Microbiol.*, vol. 6, pp. 28–40, 2008.

[183] Thattai, M. and van Oudenaarden, A., "Stochastic Gene Expression in Fluctuating Environments," *Genetics*, vol. 167, no. 1, pp. 523–530, 2004.

[184] Tian, T. and Burrage, K., "Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage lambda," *J. Theor. Biol.*, vol. 227, pp. 229–237, 2004.

[185] Veening, J., Smits, W. K., and Kulpers, O. P., "Bistability, epigenetics, and bet-hedging in bacteria," *Annu. Rev. Microbiol.*, vol. 62, pp. 193–210, 2008.

[186] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Schochet, O., "Novel type of phase transition in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, pp. 1226–1229, 1995.

[187] Vilar, J. M. G. and Saiz, L., "DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise," *Curr. Opin. Genes Dev.*, vol. 15, pp. 136–144, 2005.

[188] Wagner, A., "Robustness against mutations in genetic networks of yeast," *Nat. Genet.*, vol. 24, pp. 355–361, 2000.

[189] Wang, T. J., Gona, P., Larson, M. G., Tofler, G. H., Levy, D., Newton-Cheh, C., Jacques, P. F., Rifai, N., Selhub, J., Robins, S. J., Benjamin, E. J., D'Agostino, R. B., and Vasan, R. S., "Multiple biomarkers for the prediction of first major cardiovascular events and death," *New Eng. J. Med.*, vol. 355, no. 25, pp. 2631–2639, 2006.

[190] Webb, C. D., Graumann, P. L., Kahana, J. A., Teleman, A. A., Silver, P. A., and Losick, R., "Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in *Bacillus subtilis*," *Mol. Microbiol.*, vol. 28, no. 5, pp. 883–892, 1998.

[191] Wei, J. S., Greer, B. T., Westermann, F., Steinberg, S. M., Son, C.-G., Chen, Q.-R., Whiteford, C. C., Bilke, S., Krasnoselsky, A. L., Cenacchi, N., Catchpoole, D., Berthold, F., Schwab, M., and Khan,

J., "Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma," *Cancer Res.*, vol. 64, no. 19, pp. 6883–6891, 2004.

[192] Weinberger, L., Dar, R. D., and Simpson, M. L., "Transient-mediated fate determination in a transcriptional circuit of HIV," *Nat. Gen.*, vol. 40, pp. 466–470, 2008.

[193] Weitz, J. S., Mileyko, Y., Joh, R. I., and Voit, E. O., "Collective decisions among bacterial viruses," *Biophys. J.*, vol. 95, pp. 2673–2680, 2008.

[194] Werner, S. L., Barken, D., and Hoffmann, A., "Stimulus specificity of gene expression programs determined by temporal control of ikk activity," *Science*, vol. 309, no. 5742, pp. 1857–1861, 2005.

[195] West, S., Griffin, A., Gardner, A., and Diggle, S., "Social evolution theory for microorganisms," *Nat. Rev. Microbiol.*, vol. 4, pp. 597–607, 2006.

[196] Widmann, C., Gibson, S., Jarpe, M., and Johnson, G., "Mitogen-activated protein kinase: Conservation of a three-kinase module from yeast to human," *Physiol. Rev.*, vol. 79, no. 1, pp. 143–180, 1999.

[197] Wilson, P. W. F., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B., "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.

[198] Wong, C. C., K, E, L., Bossert, N. L., De Jonge, C. J., Baser, T. M., and Reijo Pera, R. D., "Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst state," *Nat. Biotech.*, vol. 28, pp. 1115–1121, 2010.

[199] Xiong, W. and Ferrell Jr., J. E., "A positive-feedback-based bistable memory module that governs a cell fate decision," *Nature*, vol. 426, pp. 460–465, 2003.

[200] Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R., "Myosin v walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization," *Science*, vol. 300, no. 5628, pp. 2061–2065, 2003.

[201] Zeng, L., Skinner, S. O., Zong, C., Slippy, J., Feiss, M., and Golding, I., "Decision making at a subcellular level determines the outcome of bacteriophage infection," *Cell*, vol. 141, no. 4, pp. 682–691, 2010.

[202] Zgaga, V., Medić, M., Salaj-Smic, E., Nowak, D., and Wrischer, M., "Infection of *Escherichia coli* envelop-memberne complex with lambda phage: adsoprtion and penetration," *J. Mol. Biol.*, vol. 79, no. 4, pp. 697–708, 1973.

[203] Ziv, E., Nemenman, I., and Wiggins, C. H., "Optimal signal processing in small stochastic biochemical networks," *PLoS One*, vol. 2, p. e1077, 10 2007.